

# KNOWLEDGE GRAPH DACH

MARKT SCHWABEN, 28.10.2018

**ADRESSE**

Wallbergstraße 7  
85570 Markt Schwaben  
[www.tourismuszukunft.de](http://www.tourismuszukunft.de)

**KONTAKT**

Florian Bauhuber  
Tel. +49 16099189560  
[f.bauhuber@tourismuszukunft.de](mailto:f.bauhuber@tourismuszukunft.de)

**VERTRETEN DURCH**

Tourismuszukunft Holding GmbH  
HRB 237530 | Registergericht München  
Florian Bauhuber

# INHALT

1.	Ausgangssituation .....	3
2.	Ein gemeinsamer touristischer Knowledge Graph.....	5
2.1	Warum ist jetzt die Zeit für Open Data? .....	5
2.2	Was ist zu tun? .....	5
2.3	Nutzung des globalen Standards schema.org.....	6
2.4	Erweiterung des Schema.org Vokabulars.....	7
2.5	Öffnung der Daten und Verknüpfung mit der Linked open data Cloud.....	7
2.6	Ein touristischer Knowledge Graph für den DACH-Raum .....	8

**ADRESSE**

Wallbergstraße 7  
85570 Markt Schwaben  
[www.tourismuszukunft.de](http://www.tourismuszukunft.de)

**KONTAKT**

Florian Bauhuber  
Tel. +49 16099189560  
[f.bauhuber@tourismuszukunft.de](mailto:f.bauhuber@tourismuszukunft.de)

**VERTRETEN DURCH**

Tourismuszukunft Holding GmbH  
HRB 237530 | Registergericht München  
Florian Bauhuber

## 1. AUSGANGSSITUATION

Die Digitalisierung nimmt Einfluss auf nahezu alle gesellschaftlichen Bereiche und erfasst natürlich auch die Tourismusbranche. Die Vielzahl technischer Innovationen, v.a. im Kontext der Künstlichen Intelligenz, eröffnen ungeahnte Möglichkeiten, bringen aber gleichzeitig grundlegende Veränderungen im Kommunikations- und Konsumverhalten der Gesellschaft mit sich. Dies stellt Destinationen und touristische Akteure vor Herausforderungen, welche innovative Lösungen, flexible Strukturen und gemeinschaftliches Handeln unabdingbar machen, um mit den fortschreitenden Entwicklungen und den sich rapide ändernden Rahmenbedingungen Schritt zu halten und diese für den Tourismus systematisch zu nutzen.

Eine zentrale Rolle in der Digitalisierung des Tourismus nimmt dabei die digitale Verwaltung der touristischen Inhalte (Contents) ein. Durch die aktuell hohe Menge an produzierten Inhalten befinden wir uns inmitten eines Content-Schocks: Die Content-Produktion übersteigt auch im Tourismus die Rezeptionsfähigkeit der Konsumenten um ein Vielfaches. Einfache Inhalte ohne strategische Basis und semantische Auszeichnung werden immer wirkungsloser, weil sie in einer Flut an Informationen untergehen. Nur wirklich relevante Inhalte werden noch wahrgenommen. Die Relevanz der Inhalte kann dabei jedoch nicht nur durch Quantität (mehr Inhalte) erreicht werden, sondern vor allem durch ein Mehr an Qualität.

Tourismuszukunft hat mit Change4Destination im Jahr 2016 einen Think Tank bestehend aus verschiedenen Experten der Tourismusbranche organisiert. Inhalt des Think Tanks waren strategische Fragen zum Tourismusmarketing in einer digitalen Gesellschaft. Ergebnis sind 13 Thesen zur Zukunft des öffentlich finanzierten Tourismusmarketings, welche sich unter anderem auch mit Datenstrategien im digitalen Kontext beschäftigen.

Als Kernaufgabe touristischer Akteure wird dabei der **Aufbau einer offenen digitalen Dateninfrastruktur** und die **Sicherstellung einer durchgängigen Datenqualität** und -sicherheit im jeweiligen Gebiet definiert. Diese Aufgabe wird fälschlicherweise oft an die lokalen Ebenen vor Ort delegiert, wobei dies zur Inhomogenität, zu Inkonsistenzen und offensichtlich unzureichenden Daten führt. Relevante Daten für Endkunden innerhalb einer Region (z.B. Daten zu Öffnungszeiten von POIs, Besucheranzahl in Attraktionen, Sperrungen von Wanderwegen, Badesees-Temperaturen), geschweige denn über Regions- und Ländergrenzen hinweg, sind so selten durchgängig erfasst und führen zu unnötigen Frustramenten im Urlaub. Eine gute Datenqualität zeichnet sich durch Kriterien wie Aktualität, Vollständigkeit und Verfügbarkeit aus. Eine gemeinsame Standardsprache, eine gemeinsame Auszeichnung der Inhalte bzw. die Schaffung eines gemeinsamen Wissensbasis (Knowledge Graphs) für den DACH-Raum bieten neben der Sicherstellung der Verfügbarkeit von Daten über Systemgrenzen hinweg einen Nährboden für Innovationen. Durch die vernetzte Nutzung der Daten werden neue Anreize und Mehrwerte geschaffen und damit einhergehend auch neue Produkte und Dienstleistungen.

Die digitale Dateninfrastruktur und Gewährleistung einer durchgängigen Datenqualität als Kernaufgabe von Tourismusorganisationen bzw. regional wirkender Akteure führt zu einer weiteren Konsequenz: Die Öffnung der Datensilos. Um erfolgreich als touristische Marke agieren zu können und nicht abhängig von globalen Akteuren zu werden, ist es notwendig, die Datensilos zu öffnen und mit einer offenen Datenstrategie und klaren Nutzungsregeln im Markt tätig zu werden. Offene Projekte, wie z.B. die OpenStreetMap, haben gezeigt, wie globale Monopole (GoogleMaps) konsequent aufgebrochen werden, wenn Daten für unterschiedliche Akteure verfügbar gemacht sind. Dabei geht es nicht vorrangig um ständiges Neuerfinden, sondern auch darum, auf Bestehendem aufzusetzen.

Doch was bedeutet es, seine Datenbank zu öffnen? Tim Berners-Lee, der Begründer des World Wide Web, hat dabei ein Modell vorgeschlagen, nach dem sich die **Öffnung der Daten in fünf Stufen** kategorisieren lässt („5 Stars of Linked Open Data“).

- \* = Die Daten sind unter einer offenen Lizenz online verfügbar (z.B. CC-BY), wobei das Format der Daten egal ist.
- \*\* = Die Daten stehen in einem strukturierten Format bereit (z.B. Excel statt eines eingescannten Bilds einer Tabelle).
- \*\*\* = Es werden offene, nicht proprietäre Formate verwendet, um die Daten bereitzustellen (z.B. CSV statt Excel).
- \*\*\*\* = Es werden URIs verwendet, um Dinge zu bezeichnen, damit die Daten verlinkt werden können (z.B. RDF und SPARQL).
- \*\*\*\*\* = Die Daten werden mit anderen Daten verknüpft, um einen Kontext herzustellen.

Bislang ist das Handeln aller Akteure im DACH-Raum geprägt durch eine dezentrale Content-Verwaltung und -Organisation. Eine gemeinsame (offene) Dateninfrastruktur ist nicht realisiert – egal welche Stufe man in der Skala betrachtet. Erste Test-Projekte wurden auf lokaler & Landesebene realisiert, allerdings fehlt eine übergreifende Strategie für den DACH-Raum. Dies führt zu

- einer Dopplung von Aufgaben und Zuständigkeiten,
- Inhomogenität in der Content-Qualität,
- zahlreichen Schnittstellen-Problemen und hohen Kosten,
- einer Dopplung von Inhalten in unterschiedlichen Content-Töpfen,
- zu abweichenden Content-Strukturen innerhalb unterschiedlicher Abteilungen der jeweiligen Organisationen bzw. zwischen den Organisationen.

In der Folge haben sich zahlreiche Datensilos und technische Insellösungen entwickelt. Die Digitalisierung, vor allem vor dem Hintergrund neuer semantischer Technologien, erfordert hingegen offene, einheitlich strukturierte und nachhaltig konzipierte Systemarchitekturen und Content-Strukturen.

**ADRESSE**

Wallbergstraße 7  
85570 Markt Schwaben  
www.tourismuszukunft.de

**KONTAKT**

Florian Bauhuber  
Tel. +49 16099189560  
f.bauhuber@tourismuszukunft.de

**VERTRETEN DURCH**

Tourismuszukunft Holding GmbH  
HRB 237530 | Registergericht München  
Florian Bauhuber

## 2. EIN GEMEINSAMER TOURISTISCHER KNOWLEDGE GRAPH

Viele der oben genannten Herausforderungen sind nicht neu und prägen den Tourismus bereits seit Jahren. Warum ist gerade jetzt Zeit für ein gemeinsames Handeln? Und was ist zu tun?

### 2.1 WARUM IST JETZT DIE ZEIT FÜR OPEN DATA?

- **Veränderung des Marktes:** Global Player schaffen neue Lösungen in der Kommunikation (z.B. Conversational Interfaces wie Alexa oder Ok Google) und bieten eigene Lösungen zur Kommunikation mit dem Gast entlang der Customer Journey (z.B. Facebook City Guides, Google Trips). Touristische Organisationen dekorieren (frei nach Richard David Precht) parallel „auf der Titanic die Liegestühle“ und gestalten schöne Webseiten, die immer weniger organische Aufmerksamkeit bekommen. Um in diesen Diensten mit den eigenen Daten präsent zu sein, muss man die eigenen Daten öffnen – hierfür haben sich die globalen Akteure einen gemeinsamen de facto-Standard gegeben: schema.org! Dieser Standard wird oft unbewusst von den Organisationen bereits genutzt; allerdings oft mehr schlecht als recht.
- **Politischer Rahmen:** öffentliche Organisationen im DACH-Raum werden durch den politischen Handlungsrahmen genötigt, sich mit dem Thema Open Data zu beschäftigen, wenn Steuermittel zu Erhebung der touristischen Daten eingesetzt werden. In mehreren europäischen Ländern wird schon nach der Prämisse „open by default“ gehandelt. In anderen Regionen stehen die Zeichen auf Open Data, so ist z.B. im Koalitionsvertrag der GroKo in Deutschland ist ein neues Open Data-Gesetz angekündigt.
- **Künstliche Intelligenz:** Wir stehen an einem paradigmatischen Wendepunkt unseres Handelns. Die Künstliche Intelligenz steht vor der Tür – wir müssen unsere Destinationen und Produkte für die Künstliche Intelligenz verständlich und interpretierbar machen. Der praktikabelste Weg ist es dabei, die eigenen Daten als Linked Open Data (vergleiche hierzu die 5. Stufe der 5-Sterne-Skala von Tim Berners-Lee) mit vorhandenen Knowledge Graphen im Netz (Linked Open Data Cloud) zu verknüpfen um Dezentrale Datenstrukturen zu schaffen.

### 2.2 WAS IST ZU TUN?

Im Rahmen zweier Workshops zum Thema Linked Open Data, bei dem sich zehn touristische Organisationen aus Österreich, Deutschland und Italien sowie zwei Hochschulen beteiligten, wurden gemeinsame operative Umsetzungsschritte identifiziert und erste Ergebnisse erarbeitet. Im Rahmen dieses Dokuments sollen diese technischen Schritte transparent gemacht werden, um allen Akteuren und technischen Umsetzern im deutschsprachigen Tourismus den selben Wissensstand mitgeben zu können. Niemand soll in Zukunft mehr sagen können, er habe davon nichts gewusst. Ziel ist, dass in Zukunft bei

Ausschreibungen von technischen Systemen diese Standards Anwendung finden, so dass lästige operative Probleme in der heterogenen Systemlandschaft der Vergangenheit angehören.

- a. Nutzung des globalen Standards schema.org
- b. Anpassung des Schema.org Vokabulars zur standardisierten Abbildung aller relevanten Aspekte touristischer Daten.
- c. Verknüpfung der Daten mit der Linked Open Data Cloud
- d. Schaffung eines gemeinsamen touristischen Knowledge Graphen für den DACH-Raum = DACH-KG

Neben diesen eher technischen Aspekten, wurden weitere Punkte definiert, die in Zukunft gemeinsam angegangen werden sollen:

- e. Gemeinsame Überzeugungsarbeit & die Beschaffung von Materialien zur Weiterbildung der Akteure
- f. Gemeinsamer Dialog mit den Big Playern wie Google, Facebook & Microsoft und Co.

Es erfolgte im Rahmen des Workshops keine Priorisierung der Maßnahmen. Ziel ist es in weiteren Workshops mit den Akteuren, diese Punkte voranzutreiben. Im Folgenden werden die jeweiligen Punkte detaillierter beschreiben:

### 2.3 NUTZUNG DES GLOBALEN STANDARDS SCHEMA.ORG

Wenn wir von Schema.org reden, reden wir von einem De-Facto-Standard für strukturierte Daten. Wichtige Player der Plattformökonomie (z.B. Google, Microsoft und Yandex) haben sich auf eine gemeinsame „Sprache“ für die Auszeichnung von Daten geeinigt. Diese „Sprache“ hat sich aufgrund der Marktmacht dieser Player im Markt etabliert, und wird bereits von vielen touristischen Akteuren in mehr oder weniger exakter Form für ihre Zwecke genutzt. Wir wollen auch weiterhin diesen Standard für die Auszeichnung touristischer Inhalte nutzen: allerdings einheitlich! Das heißt:

- 1.) **Mit einer durchgängigen Qualität:** die Auszeichnung für Schema.org ist nicht gleich der Auszeichnung für Schema.org! Hier existieren unterschiedliche Qualitätsniveaus im Markt. Wir wollen eine durchgängige idente Auszeichnung bei allen technischen Dienstleistern im Markt. Hierfür werden wir die Vorgaben für die Auszeichnung (=Domain Specifications) transparent machen. Alle bereits definierten Vorgaben (= Touristic Domain Specifications) findet Dienstleister unter folgendem Link: <https://semantify.it/domainSpecifications/public>

Alle beteiligten Akteure wollen in Zukunft diese Vorgaben in Ausschreibung als Pflichtbestandteil integrieren und diese Vorgaben ebenso an Partner in den jeweiligen Regionen und Destinationen distribuieren, so dass sich diese Vorgaben als Standard für die Umsetzung des Quasi-Standards in der Branche etablieren. Zudem wurde ein Plugin vom STI vorgestellt um Domain Specifications als direkt in Wordpress zu pflegen – dies soll auch für Typo 3 zur Verfügung gestellt werden, bzw. ist auch als offener Dienst (Github) für anderen CMS-Systeme zugänglich. Ziel ist es hierbei, die

strukturierten Daten so einfach wie möglich und durchgängig gleich in allen Organisationen zu pflegen – eine automatisierte inhaltliche Überprüfung ist ebenso in Arbeit.

- 2.) **Mit einer sichergestellten Qualität:** um die Auszeichnung der Dienstleister validieren zu können, wird vom Semantic Technology Institute Innsbruck (STI2) ein Tool entwickelt (<https://semantify.it/validator/>), mit dem touristische Organisationen ihre Dienstleister auf den vorgabenkonformen Einsatz der Domain Specifications überprüfen können. Dieses Tool macht für Laien klar ersichtlich, ob die Daten nach den definierten Domain Specifications ausgezeichnet sind.

## 2.4 ERWEITERUNG DES SCHEMA.ORG VOKABULARS

Für viele touristische Datentypen ist das Schema.org Vokabular noch nicht ausreichend, um die Komplexität im System Tourismus abzubilden (z.B. Skigebiete, Wandertouren, Stadtführungen). Globale Player verstehen in vielen Fällen nicht die gelebte Realität im deutschsprachigen Tourismus. Um diese Datentypen allerdings ebenso auszeichnen zu können, ist es nötig, für diese Datentypen eine gemeinsame Sprache zu entwickeln bzw. die bereits bekannte Sprache (Schema.org) so zu erweitern, dass dieses Vokabular von den Global Playern für die Erweiterung des Quasi-Standards adaptiert wird. Hierfür wollen wir allerdings keine eigenen „Wörter“ in der „Sprache“ erfinden, sondern aktiv uns bemühen, bereits von anderen Organisationen und Institutionen verwendetes Vokabular zu benutzen. Ziel ist es hier, die Welt nicht neu zu erfinden, sondern möglichst etablierte Begrifflichkeiten zu adaptieren, um den Anschluss zu anderem Vokabular (z.B. zur Mobilität oder etablierten Standards wie AlpineBits) gewährleisten zu können.

- In einem ersten Schritt werden hierfür fehlende Vokabularien in Schema.org identifiziert und priorisiert. Diese wurden im 2. Workshop der Arbeitsgruppe erarbeitet. Als besonders relevante Content-Typen, mit denen sich die Arbeitsgruppe weiter beschäftigt, wurden folgende definiert: Touren, Infrastruktur-Netz, Bikeparks, Skigebiete & Loipenparks, Erlebnisse, ÖPNV-Fahrpläne und Events. In diesen Content-Typen, aber auch in den anderen erarbeiteten Content-Typen, werden nun Ausprägungen von der Gruppe erarbeitet, die dann auf ihre Abbildbarkeit in schema.org überprüft werden.
- In einem zweiten Schritt werden dann passende „Wörter“ bzw. Domain Specifications für diese Datentypen gesucht und erarbeitet.
- Diese Domain Specifications werden in einem dritten Schritt direkt in die Vorgaben für Agenturen & Dienstleister (= Touristic Domain Specifications) überführt.
- In einem vierten Schritt werden die neuen „Wörter“ (Entities & Properties) dem Schema.org-Konsortium als Erweiterung für das Core Vokabular von Schema.org vorgeschlagen

## 2.5 ÖFFNUNG DER DATEN UND VERKNÜPFUNG MIT DER LINKED OPEN DATA CLOUD

Um die Daten nicht nur richtig auszuzeichnen, sondern sie auch sichtbar und nutzbar zu machen für alle Akteure im Internet, wollen wir die Daten mit der Linked Open Data Cloud (<https://lod-cloud.net/>) verknüpfen. Was ist die Linked Open Data Cloud? Wikipedia sagt dazu: „Linked Open Data (LOD) bezeichnet im World Wide Web frei verfügbare Daten, die per Uniform Resource Identifier (URI)

identifiziert sind und darüber direkt per HTTP abgerufen werden können und ebenfalls per URI auf andere Daten verweisen. [...] Die miteinander verknüpften Daten ergeben ein weltweites Netz, das auch als „Linked [Open] Data Cloud“ oder „Giant Global Graph“ bezeichnet wird.“ Auf uns warten also mehrere Aufgaben:

- 1.) Sicherstellung, dass überhaupt die Rechte dafür existieren, die Daten zu öffnen. Das ist explizit bei Fotos und Videos ein häufiges Problem bei touristischen Akteuren, da oft die Inhalte nicht in passender Qualität bei der Beschaffung eingekauft werden. Eine explizite Empfehlung wurde in der Gruppe der Akteure nicht erarbeitet – allerdings ist davon auszugehen, dass Daten häufiger genutzt werden, umso offener die Daten für die Nutzung sind. Public Domain oder CC0 sind die offensten Rechte-Logiken für das Zeitalter von Linked Open Data. Es wurde im Rahmen des zweiten Workshops vereinbart, vorhandene Vertragswerke auszutauschen, um Rechtsanwaltskosten zu sparen und Erfahrungen zu teilen.
- 2.) Die Auszeichnung der Daten mit URIs, explizit http-URLs, so dass sie nachgeschlagen werden können.
- 3.) Die Verlinkung der Daten mit bereits existenten Daten in der Linked Open Data Cloud.

## 2.6 EIN TOURISTISCHER KNOWLEDGE GRAPH FÜR DEN DACH-RAUM

Grundsätzlich ist es sinnvoll, dass jeder Akteur seine Daten in seiner eigenen relationalen oder graphbasierten Datenbank auszeichnet – d.h. das erweiterte Schema.org-Vokabular integriert. Dies wird von der Arbeitsgruppe auch explizit gewünscht – die jeweiligen Technikinfrastrukturen sollen weitergeführt und optimiert werden – wir wollen explizit keine Vereinheitlichung der Techniklandschaft im DACH-Raum. Um allerdings in der Linked Open Data Cloud Relevanz mit den touristischen Daten aus dem deutschsprachigen Raum zu erzeugen, macht es Sinn EINEN touristischen Knowledge Graph für den DACH-Raum zu schaffen, d.h. einen Wissensgraph in dem alle, wie oben skizziert, semantisch ausgezeichneten touristischen Informationen AN EINEM ORT zugänglich sind. Dieser große Knowledge Graph soll aus den jeweiligen dezentralen Datenbankstrukturen Inhalte aggregieren (Crawling) und in einer Graphen-Datenbank visualisieren. Da alle Akteure der Arbeitsgruppe die idente Sprache bei der Auszeichnung der Inhalte in ihren Datenbanken verwenden, ist die Zuordnung der jeweiligen Inhalte (Mapping) leicht möglich. Wir schaffen nebenbei echte Interoperabilität zwischen den technischen Systemen und Hierarchieebenen – das Ende der leidigen Schnittstellen-Thematiken! Dann braucht es auch keine technische Harmonisierung, sondern wir können unabhängig von den technischen Barrieren Daten nutzen und dadurch ihren Wert erhöhen.

Das STI hat im Rahmen des ersten Treffens der Arbeitsgruppe angeboten, den bereits in Tirol im Einsatz befindlichen Knowledge Graphen (Tiroler Knowledge Graph) für das Projekt zu erweitern und ihre Graphen-Datenbank (Ontotext) für unseren Knowledge Graph im DACH-Raum (=DACH-KG) zu nutzen. Ziel ist es, dass alle offenen touristischen Daten im DACH-Raum in diesem Knowledge Graph ausgezeichnet sind, mit relevanten anderen Graphen der Linked Open Data Cloud verlinkt und dadurch für alle Akteure im Markt uneingeschränkt nutzbar sind.



In Folge müssen für diese Inhalte klare lizenzrechtlichen Entscheidungen bei den jeweiligen Akteuren – in den jeweiligen Regionen und Institutionen - getroffen und umgesetzt werden. Die Verantwortung liegt hier bei der jeweiligen Organisation bzw. beim Urheber / Owner der jeweiligen Inhalte. Hierfür ist Überzeugungsarbeit sowie die Weiterbildung von touristischen Leistungsträgern, Redakteuren und Contentproduzenten notwendig – hierbei wollen wir zusammenarbeiten, uns gegenseitig unterstützen und gemeinsam Schulungsmaterialien produzieren.

Das Dokument versteht sich als Living Paper und wird kontinuierlich ergänzt und erweitert.

**ADRESSE**

Wallbergstraße 7  
85570 Markt Schwaben  
[www.tourismuszukunft.de](http://www.tourismuszukunft.de)

**KONTAKT**

Florian Bauhuber  
Tel. +49 16099189560  
[f.bauhuber@tourismuszukunft.de](mailto:f.bauhuber@tourismuszukunft.de)

**VERTRETEN DURCH**

Tourismuszukunft Holding GmbH  
HRB 237530 | Registergericht München  
Florian Bauhuber