

## Open Data im Deutschlandtourismus

Das Thema Open Data beschäftigt den Deutschlandtourismus und die TMBW in jüngster Zeit intensiv. Zuletzt wurde dies auf dem Destination Camp in Hamburg Ende Mai bemerkbar, als sich Vertreter von DMOs, technischen Dienstleistern, Beratungsfirmen und Forschungseinrichtungen zwei Tage intensiv mit Open Data auseinandersetzten und ein Modell für den Deutschlandtourismus entwickelten.

In diesem Zusammenhang schwirren viele Begriffe durch den Raum, angefangen bei Open Data selbst über strukturierte Daten, Creative Commons, DACH-KG, Knowledge Graph bis hin zu der kritischen Frage, ob wir nicht mehr Herr über unsere Daten sind und alles frei verfügbar zugänglich machen müssen.

Nachfolgend möchten wir die einzelnen Begriffe und Zusammenhänge erklären und das von der TMBW für Baden-Württemberg angedachte Vorgehen erläutern.

An dieser Stelle sei auch auf den Artikel „Open Data“ in der letzten Ausgabe der Tourismus Aktuell verwiesen:

[https://bw.tourismusnetzwerk.info/wp-content/uploads/2019/03/Tourismus\\_Aktuell\\_01\\_2019.pdf](https://bw.tourismusnetzwerk.info/wp-content/uploads/2019/03/Tourismus_Aktuell_01_2019.pdf) (ab Seite 16)

Begriffsdefinitionen:

### Open Data

Open Data – also offene Daten – sind alle Datenbestände, die im Interesse der Allgemeinheit (ohne Einschränkung) zur Weiterverbreitung und zur Weiterverwendung zugänglich und nutzbar gemacht sind:

*“Open data and content can be freely used, modified, and shared by anyone for any purpose” <https://opendefinition.org/>*

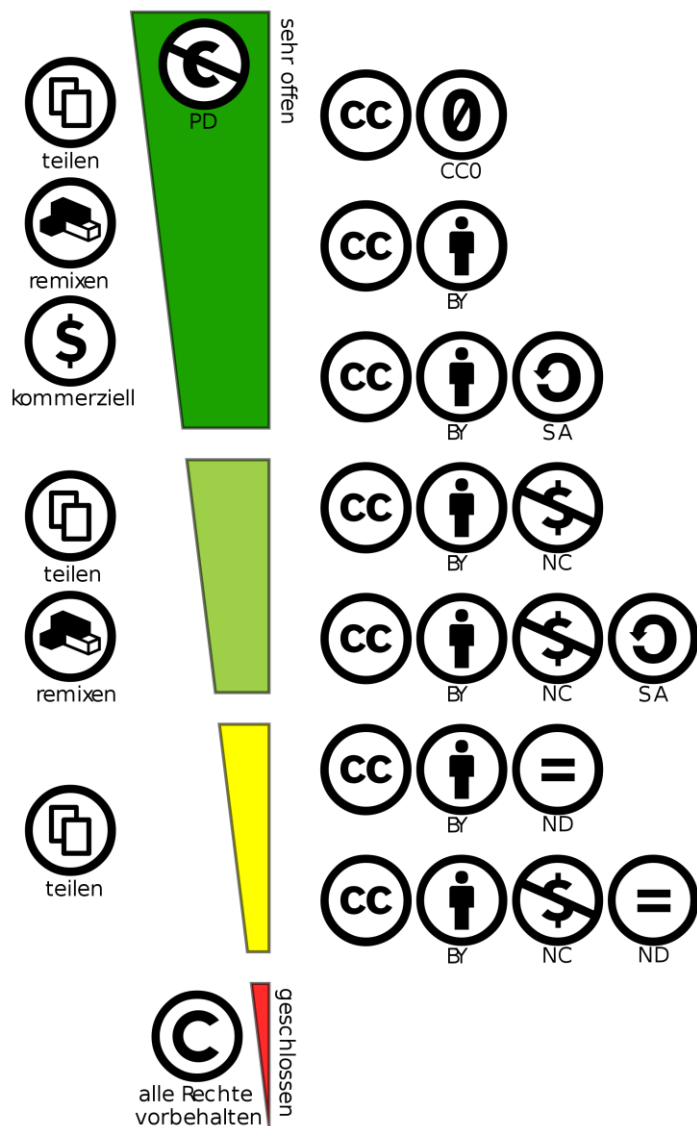
Der Begriff **Open by default** – „standardmäßig“ offene Daten – wird im Zusammenhang von Regierungen und der öffentlichen Verwaltung verwendet und bedeutet, dass öffentliche Einrichtungen und Behörden ihre Daten offen zur Verfügung stellen, um dem allgemeinen/ öffentlichen Interesse an den Daten gerecht zu werden.

Beispiel: Verwaltungsdaten, ÖPNV-Daten (keine personenbezogene Daten!)

Damit Daten offen zur Verfügung gestellt und verwendet werden können, bedarf es einer **Lizenzierung**. Auch wenn die Rede von offenen Daten ist, gibt es dennoch unterschiedliche Stufen von „offen“. Zur Lizenzierung der Daten kommen die **Creative Commons** zum Einsatz:

*„Creative Commons ist eine Non-Profit-Organisation, die standardisierte Lizenzverträge entwickelt hat, die sogenannten CC-Lizenzen. Mit ihrer Hilfe können Urheber ihre Werke gezielt und in unterschiedlichen Stufen zur Nutzung für alle freigeben.“* <https://de.creativecommons.org/>

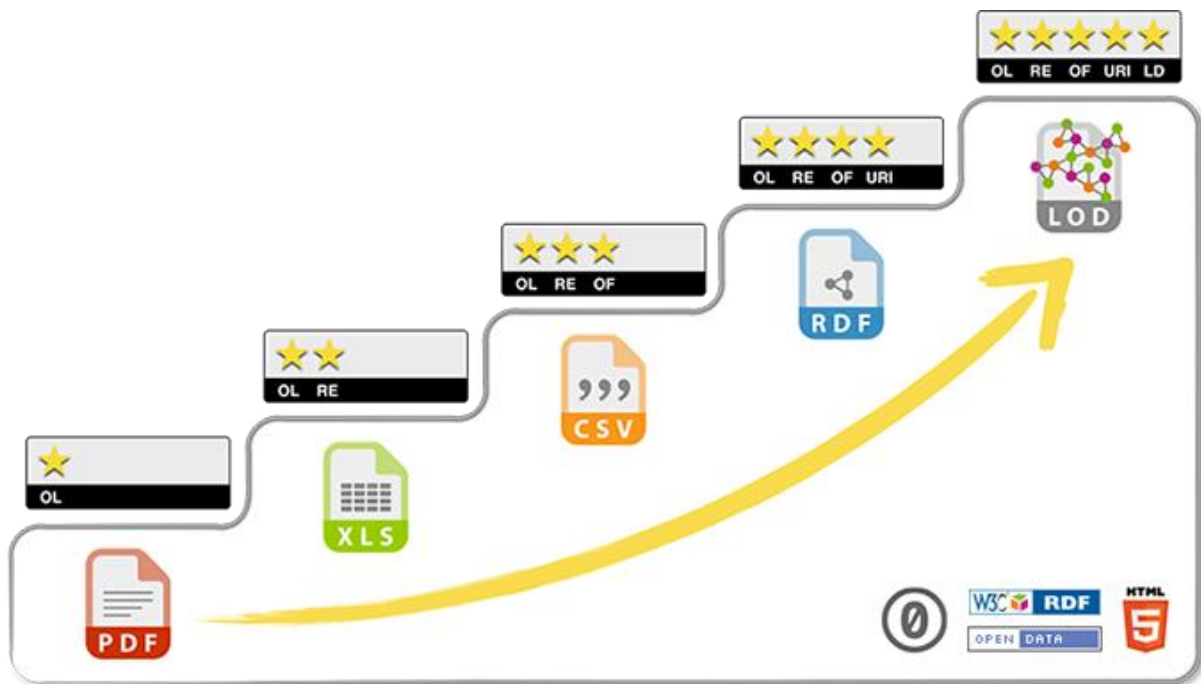
Die Creative Commons-Lizenzen reichen von sehr offen (CCO: teilen, bearbeiten, kommerziell) bis geschlossen (closed).



*Creative Commons, Quelle:*

[https://upload.wikimedia.org/wikipedia/commons/thumb/3/3d/Creative\\_Commons\\_Lizenzspektrum\\_DE.svg/2000px-Creative\\_Commons\\_Lizenzspektrum\\_DE.svg.png](https://upload.wikimedia.org/wikipedia/commons/thumb/3/3d/Creative_Commons_Lizenzspektrum_DE.svg/2000px-Creative_Commons_Lizenzspektrum_DE.svg.png)

Die höchste Stufe von offenen Daten sind nach dem 5-Sterne-Modell von Tim Berners Lee <https://5stardata.info/de/> verlinkte Daten: **Linked Open Data**. Linked Open Data sind frei verfügbare, strukturierte Daten, die dadurch mit anderen ebenfalls frei verfügbaren und strukturierten Daten semantisch verknüpft werden können.



Unter <https://5stardata.info/de/> stellt der Erfinder des World Wide Web (WWW) ein **5-Sterne-Modell** für offene Daten vor.

## Strukturierte Daten

Strukturierte Daten helfen z.B. Suchmaschinen dabei, die Mengen an Informationen, die sich auf einer Website befinden, in einen Kontext zu bringen. Hierzu werden die Inhalte mit weiterführenden Informationen verknüpft.

*„Bei der Google-Suche finden komplexe Vorgänge statt, um die Inhalte einer Seite zu erkennen. Sie können uns helfen, indem Sie strukturierte Daten in eine Seite aufnehmen und Google so explizite Hinweise zur Bedeutung der Seite geben. Strukturierte Daten sind ein standardisiertes Format, mit dem Sie Informationen zu einer Seite angeben und die Seiteninhalte klassifizieren können.“*

<https://developers.google.com/search/docs/guides/intro-structured-data?hl=de>

Bei einer Sehenswürdigkeit können dies zum Beispiel Eintrittspreise, Öffnungszeiten oder die Telefonnummer sein.

Bei der Strukturierung der Daten wird auf **schema.org** zurückgegriffen. Schema.org ist eine Initiative, die eine einheitliche Ontologie (einen Quasi-Standard) für die Strukturierung von Daten auf Websites auf der Basis von bereits bestehenden

Lizenzmodellen entwickelt. Sie wurde ursprünglich von den drei größten Suchmaschinen der Welt Google, Bing und Yahoo gebildet:

*"Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond."* <https://schema.org/>

Da schema.org nicht alle Datentypen im Tourismus abdeckt (bspw. Tourenverläufe), beschäftigt sich das Semantic Technology Institute STI mit der Erweiterung des Standards schema.org für fehlende touristische Datentypen auf Basis vorhandener touristischer Standards. Die jeweils aktuellen Informationen sind auf der Website <https://semantify.it/> zusammengestellt.

Offene Daten sind nicht gleichbedeutend mit Linked Open Data, dies fasst Prof. Dr. Eric Horster von der FH Westküste in einem Artikel zusammen: <http://www.eric-horster.de/2019/04/linked-open-data-im-tourismus-ordnung-ins-buzzword-chaos-bringen/>

### **Arbeitsgruppe DACH-KG**

Die **DACH-KG** (DACH: Deutschland, Österreich, Schweiz) ist eine Arbeitsgemeinschaft von touristischen Vertretern aus den beteiligten Ländern, die es sich zur Aufgabe gemacht hat, einen gemeinsamen touristischen Knowledge Graph für den deutschsprachigen Raum zu entwickeln.

Ausführliche Informationen zur DACH-KG:

Veröffentlichung der DACH-KG zu Definition und Zielen im November 2018:

<https://www.tourismuszukunft.de/2018/11/dach-kg-auf-dem-weg-zum-touristischen-knowledge-graph/>

Veröffentlichung der DACH-KG zu Fortschritten und Ergebnissen im Mai 2019:

<https://www.tourismuszukunft.de/2019/05/dach-kg-neue-ergebnisse-naechste-schritte-beim-thema-open-data/>

Die DACH-KG hat zum Knowledge Graph ein sog. Living Paper erstellt, das laufend aktualisiert und ergänzt wird:

<https://www.tourismuszukunft.de/wp-content/uploads/2019/04/Knowledge-Graph-DACH-V3.pdf>

### **Knowledge Graph**

Ein **Knowledge Graph** ermöglicht eine semantische – also sinnhafte und intelligente – Verknüpfung von Daten oder Content. Hier werden Dinge, Sachverhalte oder Eigenschaften in Beziehung zueinander gestellt, mit Attributen versehen und in thematische Beziehungen zueinander gebracht. Google führte bereits 2012 seinen Knowledge Graph ein, um die Suchergebnisse zu optimieren. Die Informationen werden in einer Infobox neben den Suchergebnissen dargestellt.

## Open Data im Deutschlandtourismus – deutschlandweites gemeinsames Vorgehen

Länder wie Frankreich, Schweiz, Österreich oder Irland sind beim Thema Open Data weiter voran als Deutschland.

Die DACH-KG, die Deutsche Zentrale für Tourismus (DZT) und die Landesmarketingorganisationen haben sich jeweils schon länger mit dem Thema Open Data beschäftigt. Da lag es nahe, die einzelnen Aktivitäten zusammenzuführen. Auf Initiative der DZT kamen im Herbst 2018 die Vertreter der LMOs (und teils DACH-KG) und der DZT zusammen (im Frühjahr 2019 folgten die Städte der Magic Cities), um die Herausforderungen von Open Data im Tourismus gemeinsam anzugehen und einen gemeinsamen Knowledge Graph zu schaffen.

Die Bundesländer hatten sich vorab einstimmig für ein gemeinsames Vorgehen ausgesprochen:

1. Es besteht der Wille aller Bundesländer, einen gemeinsamen Weg bei den Herausforderungen von Open Data zu gehen. Schema.org mit der Spezifikation des Semantic Technology Instituts (STI) Innsbruck wird von allen Teilnehmern zukünftig als allgemeiner Standard zur semantischen Auszeichnung touristisch relevanter Daten genutzt.
2. Die Teilnehmer unterstützen die Arbeit der DACH-Knowledge Graph Arbeitsgruppe beim STI zur Weiterentwicklung dieses Standards.
3. Alle Teilnehmer sehen die Notwendigkeit eines gemeinsamen touristischen Knowledge-Graph für Deutschland, ggf. für den deutschen Sprachraum, der sich aus dem bereitgestellten Content aller touristisch relevanten Akteure speist.
4. Dabei entscheiden die Akteure, welche Teilmengen ihres Contents den Weg in die Knowledge-Graphen finden.

Ein touristischer Knowledge Graph – eine Deutschland-Datenbank für touristischen Content – soll entwickelt werden, um den Content aus verschiedenen touristischen Datenbanken für Dritte besser zugänglich zu machen.

Dazu wurden folgende Schritte definiert:

- Alle Bundesländer erfassen ihre Daten in ihren Datenbanken nach dem Standard schema.org, im nächsten Schritt erweitert um die Standards von semantify.it
- Content-Audit durch die DZT: Ausschreibung eines Dienstleisters, der die großen Datenbanken(-anbieter) und Strukturen in Deutschland erfasst, vergleicht und Handlungsempfehlungen ableitet. Der Audit umfasst die großen Dienstleister (u.a. Land in Sicht, Infomax...). Die daraus erfolgenden Ableitungen sollen von allen Anbietern umgesetzt werden. Dienstleister, die in diesem Schritt nicht berücksichtigt wurden, sollen/ können diese Empfehlungen ebenfalls umsetzen.
- Technische Ausschreibung und Umsetzung des Knowledge Graph durch die DZT
- Geplanter Zeitrahmen für Umsetzung: bis Mai 2020

- Die Akteure in den Bundesländern erhalten Informationen und Unterstützung  
>> s. *Vorgehen in Baden-Württemberg, Hilfestellung durch TMBW*

**Ziele und Vorteile** des gemeinsamen Vorgehens:

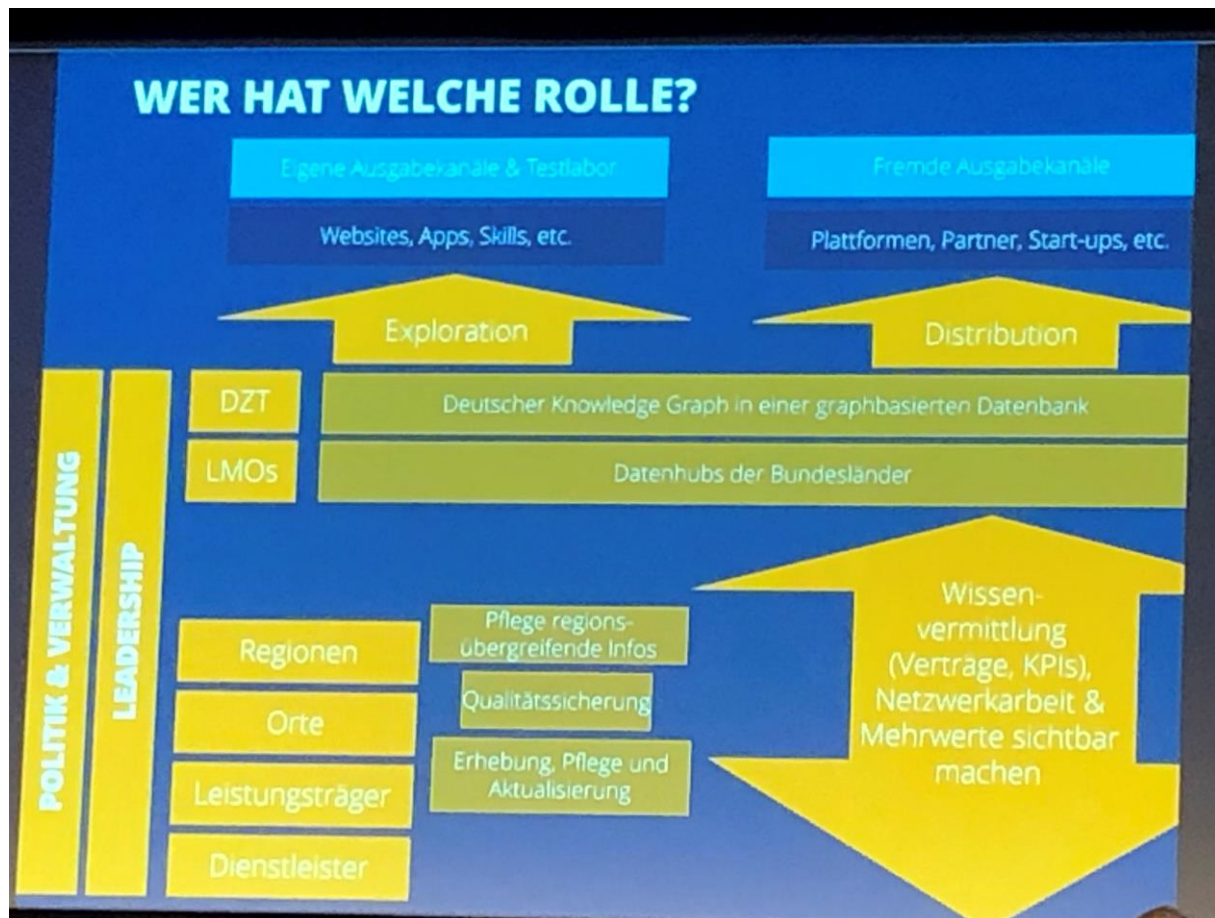
- Ende der Datensilos in und zwischen den Organisationen
- Geringere Kosten für Datenaustausch und -nutzung (Weitergabe an DZT, ADAC, DB und weitere Vertriebskanäle)
- Geringere Kosten für Beschaffung von Inhalten
- Digitaler Wandel: Ausbau der Digitalisierung im Tourismus
- Fortschreitende Künstliche Intelligenz (KI) und Machine Learning
- Bedienung von Sprachassistenten/ Conversational Interfaces/ Headless Web...
- Innovationsfähigkeit: Förderung digitaler Innovationen im Tourismus und branchenübergreifend
- Stärkung der Wettbewerbsfähigkeit: Sowohl des Tourismusstandorts Deutschlands als auch Wettbewerb innerhalb der Bundesländer
- Politischer Rahmen: über Steuermittel erhobene Daten sollen der Allgemeinheit zur Verfügung stehen
- **Für den Gast!** Die richtige Information zur richtigen Zeit im richtigen Kanal beim richtigen Gast.

Pressemitteilung Mai 2019: DZT koordiniert Open Data-Projekt zur Stärkung des Tourismusstandortes Deutschland:

[https://www.germany.travel/media/content/presse/de/PM\\_DZT\\_OpenDataStrategie\\_Projekt.pdf](https://www.germany.travel/media/content/presse/de/PM_DZT_OpenDataStrategie_Projekt.pdf)

Wie die Aufgaben- und Rollenverteilung des gemeinsamen Vorgehens aussehen könnte, wurde in einem der Workshops auf dem Destination Camp in Hamburg im Mai skizziert. Workshopteilnehmer waren unter der Moderation von Prof. Dr. Guido Sommer von der HS Kempten und Carsten Gersdorf (BestFewo) LMOs, Regionen und Touristinfos, technische Dienstleister, DZT und Berater, so dass weitgehend alle touristischen Ebenen berücksichtigt wurden.





*Open Data im Deutschlandtourismus, mögliche Rollenverteilung  
Skizze erarbeitet auf dem Destination Camp Hamburg, Mai 2019*

## Vorgehen in Baden-Württemberg – die neue toubiz-Datenbank

Baden-Württemberg schließt sich dem gemeinsamen Vorgehen auf Deutschlandebene an und unterstützt die Initiative der DZT.

Zudem ist die Umsetzung des Themas Open Data – und damit verbunden die Strukturierung und Vernetzung der Daten – im neuen Tourismuskonzept des Landes festgeschrieben. Die flächendeckende Digitalisierung der touristischen Daten bildet die Grundlage dafür, dass digitale Produkte und Services entwickelt und bedient werden können. Die TMBW möchte die Voraussetzungen und Grundlagen dafür schaffen und ausbauen sowie die Akteure im Land bei der Digitalisierung ihrer Daten weiter unterstützen.

## Datenbank toubiz®

Die Datenbank toubiz der Agentur Land in Sicht aus Freiburg wurde vor 20 Jahren eingeführt, ist bei der TMBW seit zehn Jahren implementiert und wurde über die Jahre ständig weiterentwickelt. Auch auf Regions- und Städteebene (Schwarzwald, Schwäbische Alb, Hochschwarzwald, Freiburg, Karlsruhe, Baden-Baden, Mannheim u.v.m.) sowie in vielen Orten ist die Datenbank im Einsatz und wird über mehr als 1.000 Benutzerzugänge in Baden-Württemberg gepflegt.

Wurden zu Beginn vor allem Unterkünfte und Veranstaltungen erfasst, so umfasst die Datenbank mittlerweile verschiedenste Kategorien: Veranstaltungen, Unterkünfte, Gastronomie, Direktvermarkter, Schneeberichte, Sehenswürdigkeiten/ Ausflugsziele und Touren. Bislang gibt es zwei Datenbanken: Während POIs in einer Datenbank erfasst werden, werden alle anderen Datentypen in einer separaten Maske digitalisiert. Im April 2019 waren in den Datenbanken ca. 8.700 POIs, 21.000 Veranstaltungen für die kommenden zwölf Monate und 16.000 Gastgeber hinterlegt.

Als zentrale Datenbank für Baden-Württemberg wird toubiz derzeit überarbeitet, dies umfasst folgende wesentlichen Punkte:

### *Zusammenführung der Datenbanken*

Während aktuell die POI-Datenbank über einen separaten Zugang gepflegt wird, werden die Datenbanken zusammengeführt und über einen zentralen Zugang gepflegt, so dass alle Daten an einer Stelle gepflegt werden können.

### *Open Data Check*

Mittels Fragen zu den hochgeladenen Bildern, Videos und Texten wird in toubiz ein „Open Data Check“ durchgeführt. So kann für jedes Element hinterlegt werden, wie es gemäß Creative Commons weitergegeben werden darf. Die Einordnung nach Creative Commons erfolgt in der Datenbank durch möglichst einfache Fragen, so dass sich der User selbst nicht eingehend mit den Creative Commons auseinandersetzen muss. Aus den eingegeben, strukturierten und lizenzierten Daten wird der „Open Data Pool Baden-Württemberg“ bedient, auf den zugegriffen werden kann und der auch den Knowledge Graph bedient

### *Strukturierte Daten*

Die Auszeichnung der Daten nach schema.org ist bereits in vollem Gange, die Strukturierung der Daten wird weiter umgesetzt und der Standard um semantify.it erweitert (in Abstimmung und der Veröffentlichung der Ergebnisse des STI). Die Strukturierung der Daten erfolgt im Hintergrund bzw. durch die vorgegebene Eingabemaske, der User selbst muss nichts Weiteres tun. So werden etwa Telefonnummern, Preise oder Öffnungszeiten immer automatisch einheitlich hinterlegt.

### *Geschichten und Insiderinfos*

Neben den „hard facts“ zu Ausflugszielen und interessanten Punkten soll es künftig möglich sein, Insiderwissen und besondere Tipps zu ergänzen. Viel Wissen liegt vor Ort in den Touristinfos, ist aber nicht digitalisiert. Etwa die Frage nach einem Spielplatz in der Nähe der Sehenswürdigkeit. Liegt er im Schatten? Wo ist der beste



Blick über die Stadt? Mit welcher Buslinie ist die Sehenswürdigkeit am besten zu erreichen, fährt der Bus auch an Feiertagen?

### *Tourenfassung*

Die bestehende Möglichkeit Touren in toubiz hochzuladen und zu bearbeiten wird weiter ausgebaut. So stehen u.a. die Berechnung des Höhenprofils, Downloadmöglichkeiten und die Verknüpfung zu Tourentools zur Verfügung. Auch das Frontend für die Ausgabe auf der eigenen Website wurde überarbeitet. Alternative Kartenlösungen zu Google Maps und Openstreetmap ermöglichen günstige und qualitativ hochwertige Darstellungen. Ende des Jahres werden Touren noch besser mit Wegpunkten wie POIs oder Gastgebern und aktuellen Ereignissen verknüpfbar sein.

Die grundlegende Überarbeitung der Datenbank bis Ende 2019/ Anfang 2020 zählt auf die deutschlandweite Strategie – strukturierte und geöffnete Daten – ein, und liefert damit auch die Inhalte für den Knowledge Graph, also die gemeinsame Deutschlanddatenbank.

Unabhängig vom Knowledge Graph stehen die Daten aus dem Open Data Pool Baden-Württemberg selbständig für Anwender und Vertriebskanäle zur Verfügung.

### **Vorteile:**

- Alle touristischen Daten sind an einem Ort digitalisiert, zukünftig auch Touren wie Radwege oder Wandertouren
- Kostenloses Angebot der TMBW: Die Zugänge zur Datenbank und die Pflege sind kostenlos
- Die eingetragenen Daten werden auf den TMBW-Kanälen sowie auf Orts- und Regionenebene dargestellt (z.B. Schwarzwald, Schwäbische Alb)
- Nutzung der Vertriebskanäle der TMBW: Weitergabe an ADAC, Deutsche Bahn und weitere Partner
- Aus dem Open Data Pool Baden-Württemberg können weitere Kanäle bedient werden – Datenschnittstelle
- Weitergabe der Daten an die deutschlandweite Datenbank (Knowledge Graph der DZT)
- Transparenz durch Open Data Check und spezifische Freigabe der Datensätze

### **Hilfestellung durch TMBW:**

- Für den Einkauf und die Produktion von Content wie Bildern, Videos (Auftrags- und Stockmaterial) und Texten wird die TMBW Vorlagen für Verträge erstellen lassen. Weitere Infos hierzu erfolgen in Kürze.
- Der Zugang zur Datenbank toubiz bleibt weiterhin kostenlos. Ebenso die Ausspielung der Daten auf TMBW-Ebene und im TMBW-Design.
- Informationen zur Pflege der Datenbanken werden bereitgestellt. Ebenso sollen Schulungsunterlagen erstellt werden.

Weitere Informationen werden im Tourismusnetzwerk in den kommenden Wochen und Monaten zur Verfügung gestellt.

## Häufig gestellte Fragen

### *Muss ich alle Daten freigeben?*

Nein, es besteht kein Muss die Daten freizugeben. Jedoch ist es sinnvoll, einer Freigabe zuzustimmen, da nur so Ausspielkanäle wie Regionen und TMBW, aber auch Vertriebskanäle wie DZT, ADAC oder Deutsche Bahn bedient werden können.

### *Was passiert mit meinen Daten?*

Durch den Open Data-Check in toubiz und die Freigabe der Daten (Bilder, Texte) kann genau definiert werden, welche Daten zu welchen Bedingungen frei- und weitergegeben werden. Überhaupt wird durch die Lizenzierung nach Creative Commons geregelt, welche Daten in welcher Form weitergegeben werden dürfen.

### *Verliere ich die Hoheit und Kontrolle über meine Daten?*

Sobald die Daten wie Bilder oder Texte in der Datenbank hochgeladen und mit einer offenen Lizenz versehen wurden, gehen sie ihren Weg in die Vertriebskanäle und können nicht mehr zurückgeholt werden. Auch kann nicht verfolgt oder getrackt werden, wo die Daten ausgespielt werden. Zum Zeitpunkt des Einpflegens muss klar sein, dass es für die Daten „kein Zurück mehr gibt“.

### *Warum der Knowledge Graph – können Google und Co. nicht bereits auf die Daten zugreifen?*

Durch die Strukturierung der Daten nach schema.org (und erweitert nach semantify.it) kann Google jetzt schon auf die Daten zugreifen. Ein gemeinsamer Knowledge Graph hat den Vorteil, dass die Daten über eine Stelle vermarktet werden können (z.B. über die DZT) und die Anwender der Daten nur auf eine Stelle zugreifen.

### *Unique Content und SEO – ergeben sich Nachteile beim Google Ranking?*

Welche Faktoren genau für das Google Ranking wichtig sind und wie der Algorithmus funktioniert, ist nicht bekannt und ändert sich fortlaufend. Es ist zu beachten, dass Google nur ein Kanal ist, über den die Daten ausgespielt werden. Sie stehen auch weiteren potentiellen Kanälen und Anbietern zur Verfügung.

Die Daten müssen zum Gast und nicht der Gast zu den Daten, sollte dabei die Prämisse sein.

### *Muss ich in ein bestimmtes System einpflegen?*

Im Zuge des o.g. Content-Audits der DZT werden die großen Datenbankanbieter und deren Systeme erfasst. Es sollen Standards entwickelt werden, die für die Anbieter verpflichtend sind, und auf deren Umsetzung alle gemeinsam hinarbeiten.

Weitere Anbieter können ihre Systeme entsprechend technisch anpassen. Es besteht also keine Pflicht, in ein bestimmtes System einzupflegen.

In Baden-Württemberg ist die Datenbank toubiz die zentrale touristische Datenbank fürs Land, was auch im Landestourismuskonzept festgelegt ist. Eine Pflege der Daten in toubiz wird von der TMBW gewünscht und unterstützt.

### *Entstehen Abhängigkeiten von einem Anbieter/ System?*

Es werden keine Abhängigkeiten von einem Dienstleister oder einem System entstehen, da allen Datenbanken die gleichen technischen Voraussetzungen zugrunde

liegen. Hinter der Strukturierung der Daten steht die Intention, Daten leichter von einem System in ein anderes übertragbar zu machen.

Wir freuen uns auf die Unterstützung aller und rufen dazu auf, möglichst viele Daten zu digitalisieren und so offen wie möglich zur Verfügung zu stellen. Uns ist klar, dass dies nicht von heute auf morgen passiert, sondern ein Prozess ist, der uns die nächsten Jahre begleitet. Jedoch müssen wir erkennen, dass die Zeit der Datensilos vorbei ist und wir nur gemeinsam Sichtbarkeit und Relevanz erlangen können.

**Ansprechpartnerin:**

Susanne Bleibel

Tourismus Marketing GmbH Baden-Württemberg

Digitale Kommunikation

[s.bleibel@tourismus-bw.de](mailto:s.bleibel@tourismus-bw.de)

*Stand: Juni 2019*