

# Open Data-Strategie für den brandenburgischen Tourismus

Stand: 11/2019

Lizenz: CC BY 4.0

## 1. Ausgangssituation

Die fortschreitende Digitalisierung im Tourismus bringt erhebliche Änderungen in der Produktion digitaler Inhalte (Content), aber vor allem in ihrer Distribution. War noch vor Jahren die eigene Website und deren Optimierung das Maß aller Dinge, ist es heute eher die maschinenlesbare Aufbereitung von Content das wichtigste strategische Ziel im Umgang mit Content. Dabei werden folgende Ziele verfolgt:

- Auszeichnung der eigenen Daten zur Verknüpfung mit anderen offenen Daten zur Schaffung eigener Services (semantic web, Chatbots etc.),
- Verknüpfung der eigenen Daten mit anderen touristischen Daten zur Schaffung eines touristischen knowledge Graphen zur Erzeugung von Relevanz und Nutzbarkeit für große touristische Partner und die „open data-Community“) und
- Datenbereitstellung für großen Suchmaschinen, Portale (z.B. Facebook City Guides, Google Trips) und Spracherkennungssysteme (z.B. Conversational Interfaces wie Alexa oder Ok Google)

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 Kernaufgabe touristischer Destinationen ist somit:

- der Aufbau einer offenen digitalen Dateninfrastruktur und die Sicherstellung einer durchgängigen Datenqualität und –sicherheit,
- die Auszeichnung der Nutzungsrechte für die Daten (Common Creative) und
- die Kollaboration mit möglichst vielen touristischen Partnern zu Erzeugung eines gemeinsamen offenen Graphen (semantic web).

Grundsätzlich ist die Öffnung und Kollaboration notwendig, um erfolgreich agieren zu können und nicht abhängig von globalen Akteuren zu werden. Basis der Öffnung ist eine Datenstrategie und klaren Nutzungsregeln<sup>1</sup>.

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.

Die Öffnung der eigenen Datenstruktur als „open data“ ist aber – wie oben dargestellt - nur der erste Schritt. Der zweite ist eine Vereinheitlichung der Datenbereitstellung (zumindest) im deutschsprachigen Raum, um großen Anbietern eine einheitlich zu interpretierende Datenstruktur anbieten zu können (z.B. ADAC, Deutsche Bahn).

Die Öffnung der Daten und deren „Auszeichnung“ zur Einbettung der Daten ins „semantic Web“ ist aber auch die Basis für die Integration der Daten in Spracherkennungssysteme und der Optimierung der eigenen Services durch die „Verlinkung“ der eigenen Daten mit anderen „linked open data“ (Wetter, Infrastruktur etc.).

Zudem werden öffentliche Organisationen im DACH-Raum durch den politischen Handlungsrahmen angehalten, sich mit dem Thema Open Data zu beschäftigen, wenn Steuermittel zur Erhebung der touristischen Daten eingesetzt werden. In mehreren europäischen Ländern wird schon nach der Prämisse „open by default“ gehandelt. Auch in Deutschland stehen die Zeichen auf Open Data, so ist z.B. im Koalitionsvertrag der „Großen Koalition“ in Deutschland ein neues Open Data-Gesetz angekündigt.

## **2. Umsetzungsschritte im Brandenburgischen Tourismus:**

- a. Nutzung des globalen Standards schema.org zur Repräsentation der eigenen Daten (Annotation der Daten).
- b. Integration eines Wrappers zur fortwährenden Ergänzung der schema.org Annotation
- c. Erweiterung des Schema.org Vokabulars um neue Spezifikationen, zur Abbildung aller relevanten Aspekte touristischer Daten.
- d. Definition der Nutzungsrechte und Auszeichnung der Daten
- e. Verknüpfung der Daten mit der linked Open Data Cloud
- f. Schaffung eines gemeinsamen touristischen Knowledge Graphen für den DACH-Raum
- g. Bereitstellung eines Webservice zur Abfrage der Datenbanken als JSON-LD (Entrypoint auf Basis von SPARQL)

### **a. Nutzung des globalen Standards schema.org zur Repräsentation der eigenen Daten (Annotation der Daten)**

Schema.org ist der Quasi-Standard für Auszeichnung strukturierter 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. Das heißt:

- Mit einer durchgängigen Qualität: die Auszeichnung für Schema.org wird von allen Marktteilnehmern im Tourismus unterschiedlich behandelt. Die Vorgaben für die Auszeichnung (=Domain Specifications) müssen sich an dem am meisten fortgeschrittenen Standard halten. In diesem Fall ist die Universität Innsbruck weltweit führend und hat bereits definierten Vorgaben (= Touristic Domain Specifications) erstellt. Basis sind die vorhandenen Domain Specifications nach schema.org ([www.semantify.it/domainspecifications/public](http://www.semantify.it/domainspecifications/public))

### **b. Integration eines Wrappers zur fortwährenden Ergänzung der schema.org Annotation**

Um die Daten ständig auf dem Stand der Auszeichnungsvorgaben des schema.org-Konsortium zu halten, ist ein Wrapper zu installieren, der dieses Ziel durch einen ständigen Abgleich gewährleistet.

Die nachfolgenden Schritte sind in Abstimmung mit möglichst vielen Partnern im europäischen Raum mit dem Ziel der Schaffung eines touristic knowledge graph umzusetzen.

### **c. Erweiterung des Schema.org Vokabulars um neue Spezifikationen, zur Abbildung aller relevanten Aspekte touristischer Daten**

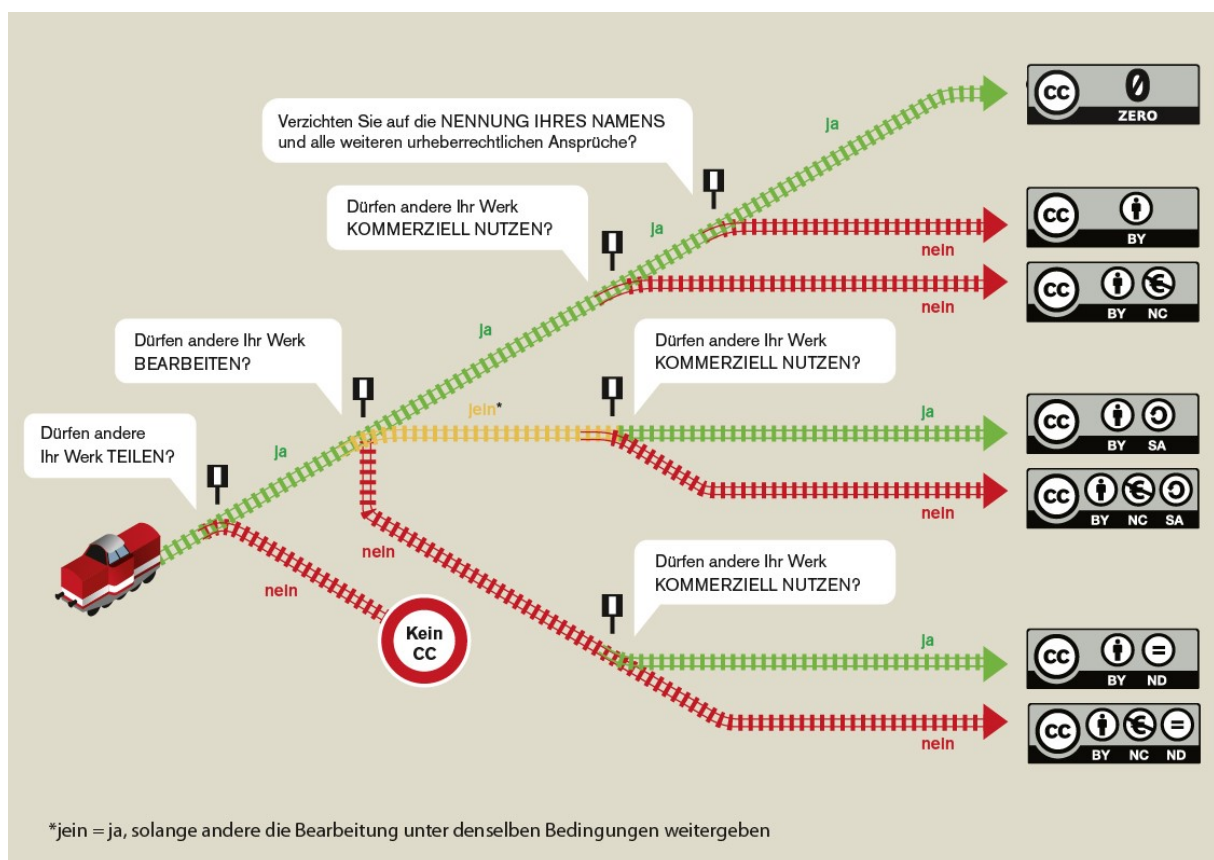
Für viele touristische Datentypen ist das Schema.org Vokabular noch nicht ausreichend, um die Komplexität im System Tourismus abzubilden (z.B. Strände, Wandertouren, Stadtführungen). Globale Player verstehen in vielen Fällen nicht die gelebte Realität im deutschsprachigen Tourismus. Um diese Datentypen 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 sollen allerdings keine eigenen „Wörter“ in der „Sprache“ erfunden werden, sondern es soll bereits von anderen Organisationen und Institutionen verwendetes Vokabular zu benutzt werden. 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 für die Beschreibung der Daten in Schema.org identifiziert und priorisiert.
- In einem zweiten Schritt werden dann passende „Wörter“ bzw. Domain Specifications für diese Datentypen gesucht und erarbeitet.
- Definition von Touristic Domain Specifications (=z.B. Entität Flughafen mit seinen Properties (Attributen), wie Airlines, Dienstleistungen, Gates etc.)
- In einem vierten Schritt werden die neuen „Wörter“ (Entities & Properties) dem Schema.org-Konsortium als Erweiterung für das Core Vokabular von Schema.org vorgeschlagen

Wie oben angedeutet, kann dieser Schritt nur zusammen mit anderen touristischen Partnern umgesetzt werden, um Relevanz gegenüber dem Schema.org-Konsortium zu erlangen. Zudem ist hier ein Dienstleister für die Umsetzung beizuholen, der wiederum das Vertrauen des Schema.org-Konsortiums genießt.

#### d. Definition der Nutzungsrechte und Auszeichnung der Daten

Die Daten des ContentNetzwerkes Brandenburg sind aber nicht nur nach schema.org zur Einbettung in das weltweite semantic web auszuzeichnen, sondern auch zur Definition der Nutzungsrechte. Hierbei muss abgestimmt werden, ob Daten einheitliche oder unterschiedliche Nutzungsrechte haben können oder welches der gemeinsame Standard ist.



Quelle: Barbara Klute und Jöran Muuß-Merholz

#### e. Verknüpfung der Daten mit der linked open data cloud

Um die Daten nicht nur richtig auszuzeichnen, sondern sie auch für alle Akteure im Internet sichtbar zu machen, werden die Daten mit der linked open data cloud (<https://lod-cloud.net/>) verknüpfen. Nach Wikipedia bedeutet LOD: „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

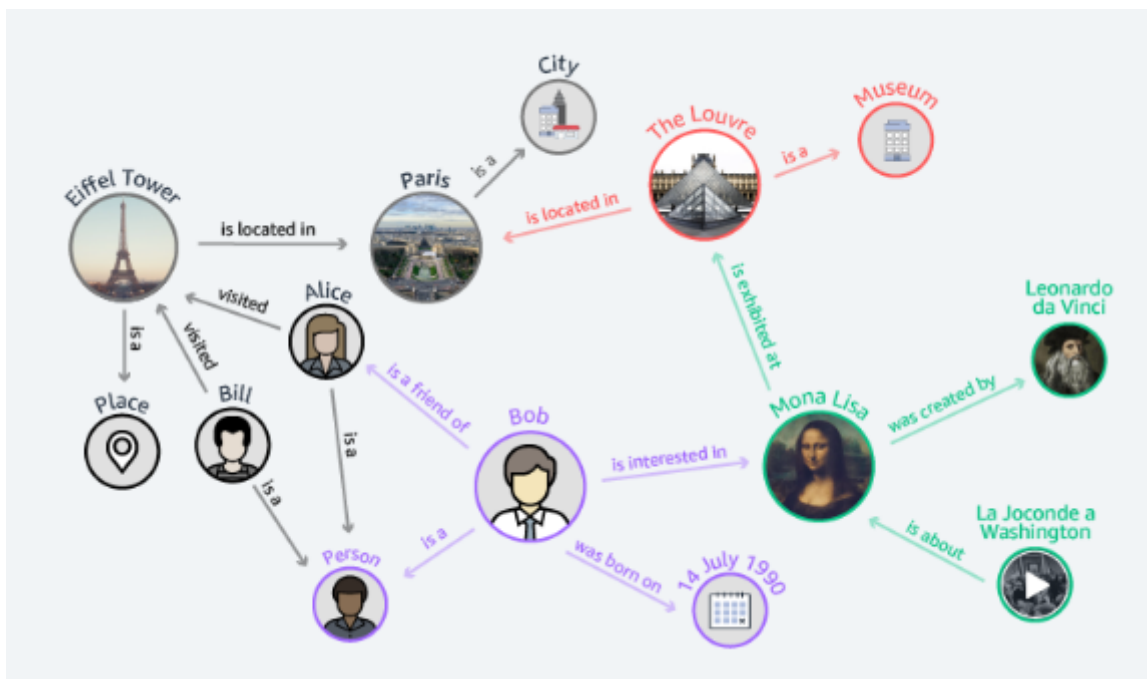
[open] data cloud“ oder „Giant Global Graph“ bezeichnet wird. Hierzu ist RDF als (Resource Description Framework) als Standard anzuwenden.

Dies impliziert folgende Aufgaben, die im Vorfeld zu lösen sind:

- Sicherstellung, dass überhaupt die Rechte dafür existieren, die Daten zu öffnen (Texte, Fotos und Videos) (s. Punkt c)
- Auszeichnung des Daten gemäß CC Standards (s. Punkt c)
- Die Verlinkung der Daten mit bereits existenten Daten in der linked open data cloud.

#### f. Schaffung eines gemeinsamen touristischen Knowledge Graphen 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. 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 zugänglich sind. Dieser große Knowledge Graph soll aus den jeweiligen dezentralen Datenbankstrukturen Inhalte aggregieren (crawling) und in einer Graphen-Datenbank visualisieren. Da alle Akteure die identische Sprache bei der Auszeichnung der Inhalte in ihren Datenbanken verwenden, ist die Zuordnung der jeweiligen Inhalte (Mapping) leicht möglich.



Wissensgraph nach <https://aws.amazon.com/de/neptune/>

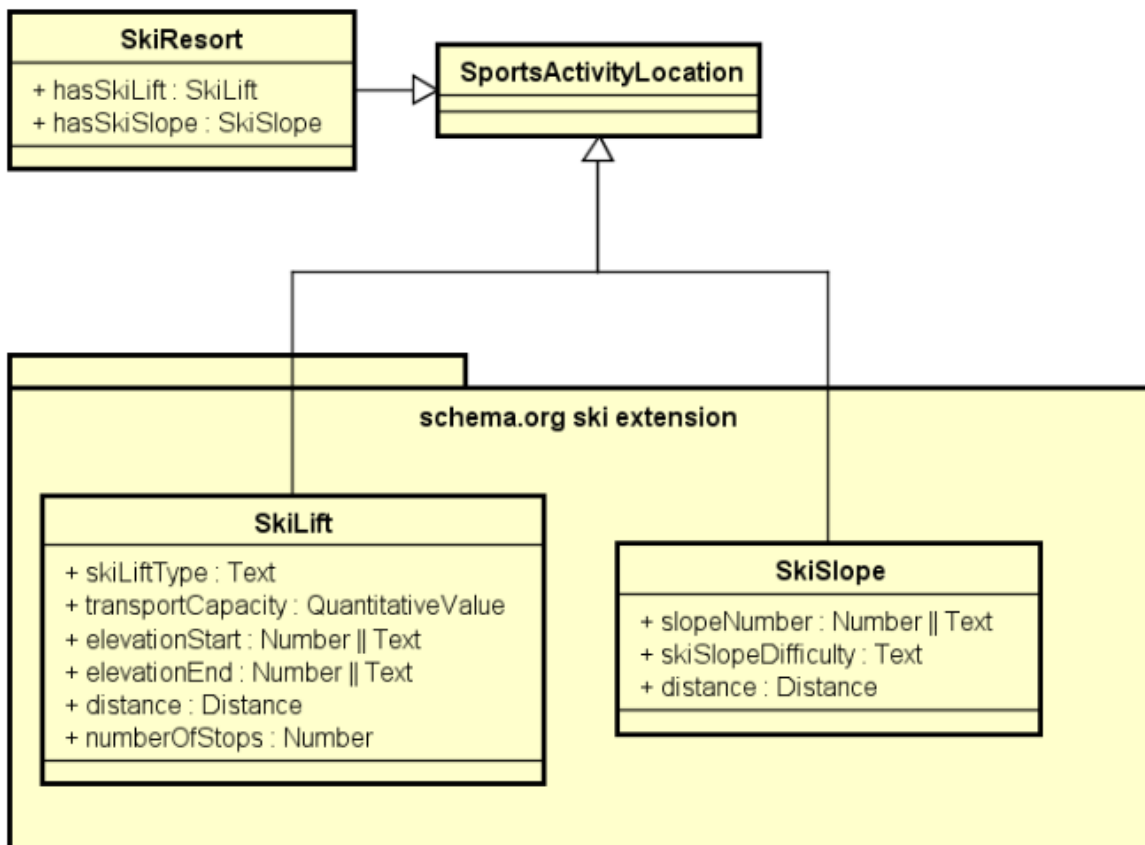
### list of new classes:

- [schema.org/SkiLift](https://sdo-skiresort-1378.appspot.com/SkiLift): <https://sdo-skiresort-1378.appspot.com/SkiLift>
- [schema.org/SkiSlope](https://sdo-skiresort-1378.appspot.com/SkiSlope): <https://sdo-skiresort-1378.appspot.com/SkiSlope>

### list of new properties:

- [schema.org/liftType](https://sdo-skiresort-1378.appspot.com/liftType): <https://sdo-skiresort-1378.appspot.com/liftType>
- [schema.org/transportCapacity](https://sdo-skiresort-1378.appspot.com/transportCapacity): <https://sdo-skiresort-1378.appspot.com/transportCapacity>
- [schema.org/elevationStart](https://sdo-skiresort-1378.appspot.com/elevationStart): <https://sdo-skiresort-1378.appspot.com/elevationStart>
- [schema.org/elevationEnd](https://sdo-skiresort-1378.appspot.com/elevationEnd): <https://sdo-skiresort-1378.appspot.com/elevationEnd>
- [schema.org/length](https://sdo-skiresort-1378.appspot.com/length): <https://sdo-skiresort-1378.appspot.com/length>
- [schema.org/numberOfStops](https://sdo-skiresort-1378.appspot.com/numberOfStops): <https://sdo-skiresort-1378.appspot.com/numberOfStops>
- [schema.org/slopeNumber](https://sdo-skiresort-1378.appspot.com/slopeNumber): <https://sdo-skiresort-1378.appspot.com/slopeNumber>
- [schema.org/difficulty](https://sdo-skiresort-1378.appspot.com/difficulty): <https://sdo-skiresort-1378.appspot.com/difficulty>
- [schema.org/hasSkiLift](https://sdo-skiresort-1378.appspot.com/hasSkiLift): <https://sdo-skiresort-1378.appspot.com/hasSkiLift>
- [schema.org/hasSkiSlope](https://sdo-skiresort-1378.appspot.com/hasSkiSlope): <https://sdo-skiresort-1378.appspot.com/hasSkiSlope>

see figure:



Beispiel für die Entwicklung einer neuen Domain Specification und Erstellung eines Vorschlages an das schema.org-Konsortium (<https://github.com/schemaorg/schemaorg/pull/1259>)

### **g. Bereitstellung eines Webservice zur Abfrage der Datenbanken als JSON-LD (Entrypoint auf Basis von SPARQL)**

Zur Bereitstellung der Daten über einen eigenen Webservice sind diese als JSON LD zur Verfügung zu stellen, der auch über SPARQL abfragbar ist.

### **3. Schrittfolge für die Umsetzung der Open Data Strategie:**

- a) Einigung auf ein gemeinsames Vorgehen zum Thema „linked open data“ im ContentNetzwerkes Brandenburg. Für das 4. Quartal 2018 ist eine erste Information im Rahmen der Erfa Digitalisierung geplant. Hier sollen Vision und Herausforderungen erstmalig dargestellt und diskutiert werden.
- b) In 2019 soll die Erweiterung der Datenbankinfrastruktur erfolgen, um die Daten nach CC Standard repräsentieren zu können. Anschließend sind in einem einjährigen Prozess die vorhandenen Medien (Bilder/Videos) zu untersuchen und auszuzeichnen.
- c) In den Jahren 2019 und 2020 werden in Zusammenarbeit mit den deutschsprachigen Partnern im Rahmen des DACH KG die „Touristic Domain Specifications“ weiterentwickelt.
- d) Im Jahr 2020-21 sollen diese touristische Annotation ins Brandenburger ContentNetzwerk übernommen werden.
- e) Auf noch zu klärenden Instanzen werden die Daten als JSON LD zur Nutzung in einem Knowledge Graphen angeboten.

Anhang:

1

Nach Tim Berners-Lee erfolgt die Öffnung der Daten in fünf Stufen („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.

Beispiel: <https://m.youtube.com/watch?v=XXK3fSH8n4k>

Stand der Entwicklung europaweit: <https://www.cappgemini.com/de-de/news/eu-studie-open-data-maturity-egovernment/>