

## 17. Wahlperiode

### Schriftliche Anfrage

#### des Abgeordneten Alexander Morlang (PIRATEN))

vom 21. August 2014 (Eingang beim Abgeordnetenhaus am 25. August 2014) und **Antwort**

#### Anzahl der Dokumenten- und Datensätze im Berliner Datenportal

Im Namen des Senats von Berlin beantworte ich Ihre Schriftliche Anfrage wie folgt:

1. Wie viele Datensätze sind aktuell im Berliner Datenportal (unter <http://daten.berlin.de/datensaetze>) verfügbar bzw. werden dort aufgelistet?

Zu 1.: Mit Stichtag 29.08.2014 werden im Berliner Datenportal 281 maschinenlesbare Datensätze verfügbar gemacht.

2. Wie viele Dokumentensätze sind aktuell im Berliner Datenportal (unter <http://daten.berlin.de/dokumente>) verfügbar bzw. werden dort aufgelistet?

Zu 2.: Mit Stichtag 29.08.2014 werden im Berliner Datenportal 490 maschinelesbare Dokumente verfügbar gemacht.

3. Nach welchen Kriterien werden „Datensätze“ und „Dokumente“ unterschieden?

Zu 3.: Für die Unterscheidung wird folgende Definition zugrunde gelegt:

##### Datensatz

Ein Datensatz fasst strukturierte Daten zu einem abgegrenzten Sachverhalt zusammen. Dabei handelt es sich oftmals um Zahlen(reihen), Codelisten oder Ähnliches in Form einer Liste, Tabelle oder Matrix. Ein Datensatz kann mit Primär- oder Sekundärdaten befüllt sein. Er kann mit einem Kopf und/oder einer Legende zur Beschreibung/Erläuterung des Inhalts versehen sein. Eine Auswertung/Visualisierung ist mit einfachen statistischen Mitteln möglich.

Zu einem Datensatz zählen immer auch eine oder mehrere digitale Ressourcen, etwa Excel-, Comma Separated Value (CSV)- oder Extended Markup Language (XML)-Dateien. Unter „Ressource“ wird auch die programmierbare Schnittstelle (application programming interface – API) eingeordnet, da über sie eine unmittelbare Anbindung an einen Datensatz hergestellt wird.

##### Dokument

Ein „Dokument“ ist eine unstrukturierte (digital vorliegende) Datei mit einer beliebigen Folge von Zeichenketten. In ihrer adressatenspezifischen Aufbereitung stellen sie Sekundärdaten dar. In Aufbau und Darstellung orientieren sie sich am menschlichen Leser. Dazu zählen insbesondere Texte, die ggf. mit Daten (in Form von Tabellen oder Grafiken) angereichert sind. Diese eingebetteten Datensätze können in den Metadaten eines Dokuments zusätzlich beschrieben werden. Um eine Maschinenlesbarkeit sicherzustellen, sind bestimmte Datenformate zu verwenden. Zur Auswertung von Dokumenten sind komplexe semantische Verfahren erforderlich. (Quelle: Abschlussbericht der Projektgruppe Open Data Berlin 2014)

4. Wie viele der Dokumenten- und Datensätze wurden jeweils von den Senatsverwaltungen (ohne nachgeordnete Behörden, Anstalten oder Eigenbetriebe) veröffentlicht?

Zu 4.: 540.

5. Wie viele der Dokumenten- und Datensätze wurden jeweils von den Bezirksverwaltungen (ohne nachgeordnete Anstalten) veröffentlicht?

Zu 5.: 23.

6. Wie bewertet der Senat diese Zahlen (s. 4. und 5.) angesichts des nunmehr fast dreijährigen Bestehens des Datenportals?

Zu 6.: Der Senat bewertet den Stand als sehr gut. Zum Vergleich: Die Stadt Wien bietet nach 3 ½ Jahren nur 235 Datensätze.

Berlin, den 09. September 2014

In Vertretung

Guido B e e r m a n n

.....  
Senatsverwaltung für Wirtschaft,  
Technologie und Forschung

(Eingang beim Abgeordnetenhaus am 16. Sep. 2014)