



Ein Jahr Open Data Portal Berlin

Erfahrungen und Ausblick

Wolfgang Both

Wolfgang Both, Senatsverwaltung für Wirtschaft, Technologie und Forschung, Berlin, wolfgang.both@senwtf.berlin.de

Abstract: Das Berliner Datenportal ist jetzt ein Jahr online. Im Probebetrieb konnten umfangreiche Erfahrungen zum Nutzinteresse und zu internen Abläufen bei der Datenerfassung, -aufbereitung und -bereitstellung gewonnen werden. Gegenwärtig läuft die Überführung des Angebots in den Regelbetrieb.

Keywords: Open Government, Open Government Data, Implementation

Das Bundesland Berlin hat sich frühzeitig auf die Öffnung seiner Datenbestände (Open Data) vorbereitet. Bereits im Frühjahr 2010 wurden gemeinsam mit dem Fraunhofer-Institut FOKUS Arbeiten an einer Vorstudie aufgenommen. In diesem Rahmen wurde untersucht, wie die heute dezentral vorhandenen Datenbestände zugänglich und verfügbar gemacht werden können. Ein erster Vorschlag für eine IT-Architektur wurde um eine inhaltliche Kategorisierung der Datenbestände ergänzt. Im Ergebnis wurden drei Vorschläge unterbreitet:

- politisches Bekenntnis zu Open Government/Open Data
- technische Umsetzung einer Daten-Cloud
- Vorbereitung durch kleinere Erprobungsvorhaben

Mit einem Beschluss des Berliner Staatssekretärsausschusses für die Verwaltungsmodernisierung im Frühjahr 2011 wurde den ersten beiden Empfehlungen umgehend entsprochen.

Dem waren zwei weitere Aktivitäten zur Sensibilisierung und Ansprache der Öffentlichkeit vorausgegangen: Zum einen startete im Herbst 2010 mit „Apps4Berlin“ der erste App-Wettbewerb einer deutschen Stadt. Zum anderen fand ein Online-Voting zu interessierenden Datenkategorien statt. Teilnehmerzahl (mehr als 1.500 Votes) und Ergebnis bekräftigten das Interesse der Berlinerinnen und Berliner an einer breiten und strukturierten Datenbereitstellung.

1. Projektgegenstand

Unter dem programmatischen Titel „Von der öffentlichen zur offenen Verwaltung“ wurde die Projektarbeit aufgenommen. Gemäß Auftrag gliederte sich das Vorhaben in sechs Arbeitspakete. Neben dem Projektmanagement wurden die folgenden Fragestellungen aufgegriffen:

- Bedarfs- und Potenzialanalyse
- Organisationsanalyse

- rechtliche Rahmenbedingungen
- Geschäftsmodelle
- Plattformkonzept, Technik und Standards

Im Ergebnis lagen Ende 2011 neben den Ausarbeitungen zu diesen Themen 39 Handlungsempfehlungen an Politik, Verwaltung, Recht und Technik vor, die in eine kurz-, mittel- und langfristige Perspektive gruppiert sind (<http://bln.io/bEw5i3q>).

Das Motto für Berlin lautet also: Transparenz, Partizipation und Effizienz. Mit der Öffnung der Datenbestände wird ein erster wichtiger Schritt in Richtung transparentes Regierungshandeln gemacht. Die Landesregierung sieht in diesem Angebot an die Bürger, die Unternehmen, die Forschung und die Medien einen Beitrag zur Weiterentwicklung unseres demokratischen Gemeinwesens. Der berechtigten Erwartung nach Teilhabe kann hier auf qualifizierter Ebene entsprochen werden.

2. Datenportal

Mit dem Aufbau und dem Internetangebot eines Datenportals wurde ein zentraler Zugang zu den bisher verteilt vorliegenden Datenbeständen der Landesverwaltung geschaffen. Prinzip ist dabei, dass die Originaldaten bei den Bearbeitern verbleiben, sie dort weiterhin gepflegt und aktualisiert werden. Damit bietet das Datenportal eine gut strukturierte Vermittlung zwischen Datennutzer und Datenquelle.

Um dies einheitlich zu gestalten, die Daten gleichermaßen zu beschreiben und eine Vergleichbarkeit zu ermöglichen, wurde im Projekt viel Aufwand in den Entwurf eines Metadatenschemas gesteckt. Die Metadaten beschreiben ein einzelnes Datum, einen Datensatz, hinsichtlich Titel, Quelle und Format, enthalten des Weiteren Angaben zur räumlichen und zeitlichen Abdeckung, zum Bearbeiter oder zu den Nutzungsbedingungen. Das Metadatenschema entstand in gründlicher Auswertung bestehender Angebote, z.B. der Datenportale von London oder Wien. Es baut auf dem Dublin-Core auf, einer Sammlung von Konventionen zur Beschreibung von Objekten und Dokumenten im Netz. Dadurch werden Objekte leichter auffindbar, vergleichbar und verknüpfbar.

Die Ablage der Metadaten (nicht des Datums selber) erfolgt in einem Datenregister. Dies stellt das Backend des Datenportals dar. Im Rahmen der Vorbereitungen für das Berliner Datenportal haben wir uns für das Comprehensive Knowledge Archive Network (CKAN) der Open Knowledge Foundation, einem Open Source-Paket, entschieden. Zum einen erlaubte uns dies eine schnelle Realisierung, zum anderen wurde CKAN bereits in anderen Städten eingesetzt. Ein Leistungsmerkmal von CKAN ist die Föderation, das heißt, die Möglichkeit der Verknüpfung unter- oder übergeordneter Aggregationsebenen von Datenbeständen. Somit können Berliner Inhalte direkt auch in übergeordneten Ebenen, wie Bund oder Europa, bereitgestellt werden.

Das Schaufenster nach außen ist das Datenportal (<http://daten.berlin.de>). Die Oberfläche ist schnell und übersichtlich mit einem Drupal Content Management System gestaltet worden. Hier geht es weniger um aufwändige Gestaltung und Navigation. Denn die Datenbestände sind vorwiegend maschinenlesbar gestaltet. Vielmehr muss eine Suche schnell zum Ergebnis führen und es Maschinen ermöglichen, sich einfach durch den Datenbestand zu arbeiten. Daher ist als ein weiteres Merkmal in den Metadaten eine Datenkategorie vorgegeben. Diese Zuweisung beschleunigt die gezielte Suche. Gleichzeitig hilft sie dem menschlichen Besucher der Seite bei der Navigation durch das Angebot.

Registrierte Datenbearbeiter können selbstständig neue Datensätze einstellen bzw. bestehende aktualisieren. Zum einen ist ein manueller Eintrag in das Datenregister möglich (hierfür liegt ein Leitfaden vor). Zum anderen bietet das CMS-Imperia-Werkzeug „SimpleSearch“ eine Exportfunktion in das CKAN-Register, mit der alle Merkmale zu den Metadaten gleich erfasst und übertragen werden. Bei Schulungen wird zukünftig stärker auf diese Funktionalität hingewiesen.

Ummittelbar nach der Berlin-Wahl im September 2011 standen bereits die Wahlergebnisse maschinenlesbar als neuer Datensatz im Portal, bereitgestellt vom Amt für Statistik Berlin-Brandenburg. Nur eine Woche später war die „Wahl-App“ fertig, die Wahlergebnisse und Bevölkerungsdaten miteinander verknüpfte.

3. Ergebnisse

Obwohl das Angebot von der Denkmalliste über die Konjunkturdaten bis hin zu Wanderwegen im Wesentlichen aus maschinenlesbaren Rohdaten besteht, verzeichnen wir einen stetigen Besucherstrom auf unser Datenportal. Nach dem ersten Hype im September hat sich die Nutzerzahl auf werktäglich ca. 1.000 Besucher eingependelt. Der Tagesrhythmus folgt dem Sonnenstand, mittags sind also die meisten Besucher auf der Plattform. Vorrangig besucht werden die Datensätze. Da die Klickrate auf die Datensätze inzwischen höher ist als auf die Startseite, werden die Datensätze offenbar mehr und mehr von Suchmaschinen direkt gefunden.

Der Datenbestand ist kontinuierlich gewachsen, aus Landes- und Bezirksverwaltungen kommen stetig neue Beiträge. Gegenwärtig sind fast 100 Datensätze im Register eingetragen, die bis zu 20 einzelne Datensätze, wie Wander- oder Radwege in einem Geodatenformat beinhalten können.

4. Ausblick

Da Weiterentwicklung und Ausbau des Angebots bisher noch Projektstatus hatten, wurde im Juni 2012 durch den Ausschuss für die Verwaltungsmodernisierung beschlossen, eine ressortübergreifende Arbeitsgruppe einzusetzen. Sie hat sich im August konstituiert, soll sich offenen Fragen der Harmonisierung rund um die Datenbeschreibung und Bereitstellung widmen und ein entsprechendes Weiterbildungsangebot für die Mitarbeiterinnen und Mitarbeiter entwickeln.

Ein wichtiges Teilprojekt ist die Entwicklung und der Einsatz eines Metadaten-Harvesters, um aus bestehenden Datenbeständen (Geodaten, Gesundheitsdaten, Sozialdaten, Demografiedaten usw.) eine Sichtbarmachung im Open Data-Portal zu erreichen. Derartige Software-Werkzeuge werden bereits vielfach eingesetzt (Bibliothekswesen), um Spezialbestände allgemein verfügbar zu machen. Manuelle Erfassung und Übertragung werden durch eine Softwarelösung automatisiert. Aus einem Register kann in ein anderes transferiert werden, so dass aufbereitete Datenbestände schneller zugänglich und nutzbar gemacht werden können.

Offene Fragen gibt es vor allem im rechtlichen und technischen Umfeld der Öffnung und des freien Zugangs.

So wurde in der Berliner Studie festgestellt, dass das rechtliche Umfeld durch viele Fachgesetze recht unübersichtlich und nicht harmonisch gestaltet ist. Daher ist in Berlin geplant, den Punkt der Datenbereitstellung in einem übergreifenden e-Government- und Organisationsgesetz mit zu regeln. Auch der Bund arbeitet an einem entsprechenden Gesetz.

Einer weiteren Klärung bedürfen die Lizenz- und Nutzungsbedingungen. Wir haben uns erst einmal an der Creative Commons Licence, die auch in Wikipedia für dort eingestellte Inhalte angewendet wird, orientiert. Diese Lizenzform ist im Internet weitgehend bekannt und akzeptiert, andererseits nicht speziell für Daten ausgelegt. Hier wird zukünftig eine Weiterentwicklung und globale Harmonisierung erforderlich sein. Diese Fragen werden gegenwärtig in einer Bund-Länderarbeitsgruppe unter Federführung des Bundesinnenministeriums bearbeitet. Ein abgestimmter Vorschlag ist an den IT-Planungsrat gegangen und soll helfen, Insellösungen zu vermeiden.

Über den Autor:

Wolfgang Both ist promovierter Informationstechniker (Jg. 1950) und arbeitet seit mehr als 15 Jahren in der Senatsverwaltung für Wirtschaft, Technologie und Forschung des Landes Berlin. Er ist dort im Fachreferat für die Informations- und Medienwirtschaft zuständig für Netzwerke, Innovationen und deren Förderung. Nach seinem Studium war er in der Industrieforschung tätig und arbeitete Anfang der 90er Jahre bei einem Projektträger des Bundesforschungsministeriums.

Zu seinen Schwerpunktthemen bei der Umsetzung der länderübergreifenden Innovationsstrategie im Zukunftsfeld IKT/Medien zählen neben Open Government auch Sicherheit mit IT sowie das „Internet der Dienste“.