

Nils Barnickel, Wolfgang Both, Matthias Flüge, Ina Schieferdecker

## Städtische Data Cloud – Offene Daten für neue, innovative Unternehmen in Städten

### *City Data Cloud – Open Data for New, Innovative Businesses in Cities*

*Mit der zunehmenden Durchdringung von Informationstechnik in die verschiedensten Bereiche der Wirtschaft und Gesellschaft treten Informationen als Gut in das Zentrum der Betrachtung. Daten und Informationen sind als Rohstoffe bzw. Ressourcen einer Informationsgesellschaft zu verstehen.*

*Das Ziel der **Berlin City Cloud** ist es, (städtischen) Unternehmen, Organisationen und Bürgern der Stadt einen vertrauenswürdigen Zugriff auf städtische und öffentliche Daten zur gemeinsamen Gestaltung der städtischen Prozesse und Abläufe zu bieten. Es werden kommerzielle und öffentliche, aktuelle, kontextsensitive Daten über städtische Infrastrukturen und Ressourcen einer Stadt als auch akkumulierte Informationen bereitgestellt, um aktuelle Situationen in der Stadt transparent zu machen, Entscheidungen zu ermöglichen bzw. nachvollziehbar zu machen.*

*With the increasing penetration of information technologies into various domains of economy and society, information gets into the centre of consideration as commodity. Data and information are to be understood as raw materials and resources of an information society.*

*The objectives of the Berlin City Cloud are to provide enterprises, organisations and citizens a trustworthy access to urban and public data for the joint design of urban processes and operations. Commercial and public, current, and context-sensitive data together with accumulated information about urban infrastructures and resources in the city are made available making current situations in the city transparent, facilitating decisions and making them comprehensible.*

Stichworte: Smart Stadt, Öffentliche Daten, Offene Daten, Linked Data, City Cloud

Keywords: Smart City, Public Sector Information, Open Data, Linked Data, City Cloud

## 1. Einführung

Das Konzept zur Berlin City Cloud entstand aus der Anerkennung der Relevanz von Daten und Informationen für eine Stadt als Dienstleister für Bürger, Unternehmen, Institutionen und Gäste. Es geht insbesondere um die Erhöhung der Zusammenarbeit, als auch Transparenz und Partizipation in der Stadt. Das Konzept wird gestützt von zwei Initiativen zur Öffnung von Daten beliebiger Couleur – **OpenData** – und zur Verfügbarmachung von öffentlichen Daten – **Public Sector Information**.

Fraunhofer FOKUS hat zum Stand der öffentlichen Daten in Berlin mit der Senatverwaltung für Wirtschaft, Technologie und Frauen eine **Vorstudie** erarbeitet, nach der zwar relativ umfangreiche städtische Daten inklusive Such- und Spezifizierungsfunktionen in Berlin zur Verfügung stehen, diese jedoch statische Informationsquellen mit eingeschränkten Möglichkeiten zur Weiterverarbeitung sind, mehrheitlich in PDF (Portable Document Format) angeboten werden und oftmals akkumulierte, bewertete Daten darstellen. Bis auf diversen Statistiken fehlt es in Berlin an Rohdaten, die in wohldefinierten Formaten und an offenen Schnittstellen für die Weiterverarbeitung angeboten werden.

Zudem wurde in Zusammenarbeit mit berlin.de eine **Umfrage** an die Berliner und Brandenburger durchgeführt, die ein Interesse an offenen Daten insbesondere zur Stadtentwicklung, Verwaltung und Umwelt ergibt.

Die (städtischen und stadtnahen) Unternehmen und Bürger sind aktiv bei der Bereitstellung und Nutzung der Daten und Informationen beteiligt: die Berlin City Cloud ermöglicht eine effiziente Bereitstellung, Analyse und Weiterverarbeitung der städtischen Daten in Diensten und Applikationen, so dass eine offene Atmosphäre zur Arbeit zu und mit den Daten und Informationen im Interesse der Stadt entsteht.

## 1.1 Stand öffentlicher Daten in der Welt

Die PSI-Richtlinie der EU die Nutzung und Verwertung der Informationen des öffentlichen Sektors stimulieren. Die Praxis zeigt aber, dass die Umsetzung dieser Anregung durch viele Hürden behindert wird. Eine Studie im Auftrag der EU ermittelte 2009 [3][9] für drei Marktsegmente (Geodaten, juristische Daten, Wetterdaten), dass sowohl restriktive Lizenzmodelle und hohe Preise sowie unterschiedliche Sichtweisen von öffentlicher Hand und Unternehmen bisher zu keinen nennenswerten Geschäftsbeziehungen geführt haben. Die Restriktionen haben vielmehr zu einer Substitution von öffentlichen Datensammlungen durch private Datenanbieter geführt (Adress- und Firmenlisten, private Wetterdienste).

Mit dem Amtsantritt von Barack Obama in den USA wurde am 21. Jan. 2009 ein Memorandum für Transparenz und offenes Regierungshandeln herausgegeben. Im Sommer 2009 wurde die Internetseite data.gov freigeschaltet, die etwa 1.000 Datensätze umfasst. Am 8. Dez. 2009 wurde das Memorandum durch die „Open Government Richtlinie“ untersetzt und alle Regierungsstellen um Vorschläge zur Umsetzung gebeten. Im Mittelpunkt stehen Transparenz, Partizipation und Zusammenarbeit. Entsprechende Zulieferungen liegen seit Mai 2010 vor.

Zahlreiche Städte, besonders in Nordamerika, haben ihre Datenbestände für die Öffentlichkeit geöffnet und elektronisch zur Verfügung gestellt (natürlich werden auch dort keine personenbezogenen oder sicherheitsrelevanten Daten freigegeben). U.a. unterstützt die Fa. Microsoft die OpenData-Aktivitäten in den USA und Kanada. So hat Vancouver zu den Olympischen Spielen einen Datenkatalog veröffentlicht. Dieser ist allerdings nur alphabetisch geordnet, es gibt keine Sachklassifikation. Ein gleichartiges Angebot liegt z.B. von Columbia vor. Von Vorteil ist, dass Datenformate und Georeferenzierungen (auf Google Maps oder Bing) mit angegeben werden.

In Europa macht sich insbesondere Großbritannien für Open Data stark. Ende Januar 2010 hat die britische Regierung die Internetseite data.gov.uk eingerichtet und ca. 3.000 Datensätze

freigeschaltet. Greater London hat eine eigene Seite [data.london.gov.uk](http://data.london.gov.uk) eingerichtet. Seit Frühjahr sind bisher 260 Datensätze über die Stadt in 17 Kategorien verfügbar.

Daneben sind eine Europa-übergreifende und zahlreiche nationale OpenData-Netzwerke entstanden. In Deutschland ist die Gruppe [opendata-network.org](http://opendata-network.org) aktiv, in Österreich ist es [opendata.at](http://opendata.at). In Deutschland ist bislang nur aus München und Stuttgart bekannt, dass es erste Überlegungen von Kommunen zu Open Data gibt. Nicht bekannt ist dagegen, wie die bereitgestellten Daten abgefragt werden und tatsächlich zu Zusammenarbeit, Partizipation und neuen Geschäften anregen.

## 2. Bestandsaufnahme der öffentlichen Datenbestände in Berlin

Die Abbildung 1 enthält eine Kategorisierung städtischer Daten. Diese werden unterteilt nach ihrer Herkunft und ihrem inhaltlichen Bezug, ihren technischen und rechtlichen Eigenschaften und ihrer Verfügbarkeit einschließlich Zugriffsmöglichkeiten, Form und Kosten.

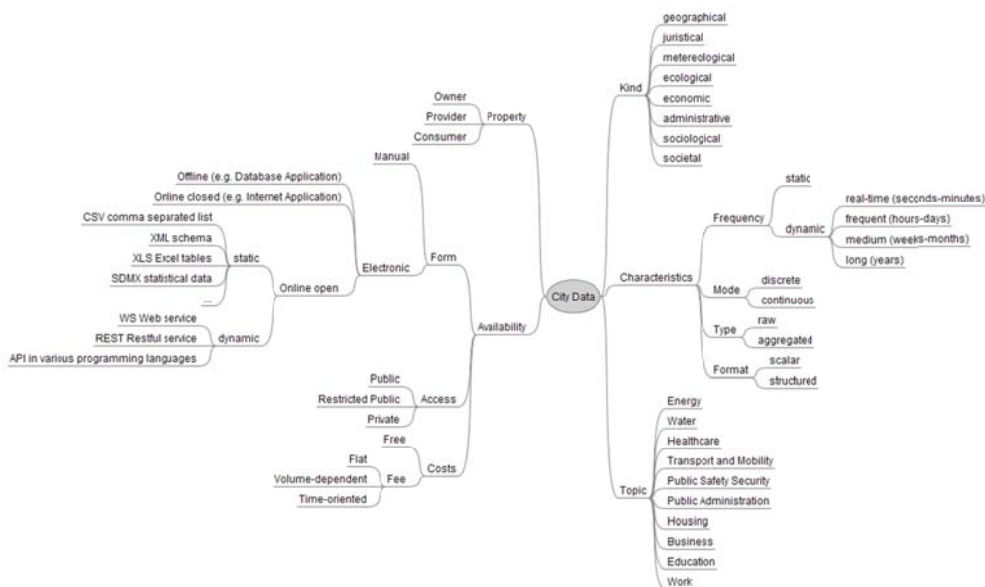


Abbildung 1 Kategorisierung öffentlicher Daten

Bei den Inhalten werden beispielsweise Wirtschafts-, Verkehrs-, Umwelt- und Klima- oder Ver- und Entsorgungsinformationen unterschieden. Die technischen Eigenschaften umfassen beispielsweise die Rate der Datenaktualisierung diskreter Daten oder die Struktur der Daten. Zu den rechtlichen Eigenschaften gehören beispielsweise die Eigentums- und Nutzungsverhältnisse. Schlussendlich sind ebenso die Verfügbarkeit in digitaler Form und die Kosten der Datennutzung zu betrachten.

Wesentliche Eigenschaften bei der Bereitstellung städtischer Daten sind, dass diese in von Computer verarbeitbaren, wohldefinierten Formaten zur Verfügung gestellt werden. Dazu zählen unter anderem:

- CSV (Character Separated Values) Formate, ein für einfache Datenstrukturen gebräuchliches Format

- XML (eXtended Markup Language) Dateien, deren Strukturen typischerweise in XML Schemata definiert sind
- SDMX (Statistical Data and Metadata Exchange) als Format für statistische Daten
- DXF (Drawing Interchange Format), SHP (Shapefile) oder DWG als Formate für Geodaten
- ECW (Enhanced Compression Wavelet) oder MrSID (Multi Resolution Seamless Image Database) als Formate für Rastergrafiken
- KML (Keyhole Markup Language) eine Auszeichnungssprache zur Beschreibung von Geodaten

## 2.1 Bewertung der Datenbestände

Die online verfügbare Vorstudie [13] enthält eine initiale Erfassung verfügbarer öffentlicher Daten in Berlin. Es wurden über 100 Datensätze identifiziert. Dabei werden die Daten in verschiedene Bereiche städtischen Lebens unterteilt:

- Wirtschaftsinformationen
- Transport- und Verkehrsinformationen
- Umwelt- und Klimainformationen
- Ver- und Entsorgungsinformationen
- Geographische Informationen
- Informationen über Medizin und Lebensmittelssicherheit
- Informationen zur öffentlichen Sicherheit
- Informationen zu Immobilien und Wohnsituation
- Informationen zu Aus- und Weiterbildung
- Informationen zu Verwaltung und Steuern
- Informationen zum Rechtssystem
- Bevölkerungsdaten
- Informationen zum Arbeitsmarkt
- Informationen zu Wahlen
- Informationen über öffentliche Sozialleistungen
- Landwirtschaftliche Produktion
- Volkswirtschaftliche Gesamtrechnungen

Es stehen demnach in Berlin bereits umfangreiche städtische Daten zur Verfügung. Sie sind vielfach mit Such- und Spezifizierungsfunktionen verbunden (Informationsschichten auf dem Stadtplan zu Hotels, Restaurants etc., Suchfunktionen zu Kindertagesstätten oder Schulen in der näheren Umgebung, Routenplaner der Berliner Verkehrs Betriebe BVG).

Jedoch sind drei Nachteile deutlich, die die Daten nur als statische Informationsquellen mit eingeschränkten Möglichkeiten zur Weiterverarbeitung nutzen lassen:

- Die Daten werden mehrheitlich in PDF (Portable Document Format) angeboten – die in eingangs genannten Formate werden kaum unterstützt.
- Die Daten sind oftmals bereits akkumulierte, bewertete Daten – bis auf diverse Statistiken fehlen die Rohdaten.
- Die Daten sind in der Mehrzahl Verlaufsdaten – aktuelle Daten zu Situationen in der Stadt sind auszubauen.

### 3. Das Konzept einer City Data Cloud

Ausgangspunkt für die Konzeption und Realisierung der City Data Cloud Berlin sind die Anforderungen aus Sicht der Nutzer, die gleichzeitig Konsumenten und Produzenten städtischer Daten sein können. Die Anforderungen sind vor einem konkreten Realisierungsprojekt detailliert aufzunehmen und zu spezifizieren. Sie orientieren sich jedoch stark am antizipierten Nutzen der involvierten Akteure.

Der Kundennutzen der City Data Cloud ist in Abhängigkeit von der betrachteten Zielgruppe recht unterschiedlich. So stehen z.B. für Bürger, Touristen und sonstige Privatkunden Informationen über städtischen Auslastungen, Bürgerbelange etc. im Vordergrund, so dass hier z.B. der einfache und endgeräteunabhängige Zugang zu Informationen als eine Anforderung zu sehen ist.

Für Unternehmen liegt der Nutzen eher in der Generierung neuer Informationsprodukte und -dienste durch die Bereitstellung eigener Daten an Andere und durch die Nutzung der Daten Anderer. So sind die Anforderungen in diesem Fall eher von Bedarfen im Bereich der Datensicherheit und der rechtssicheren Verwendung von Daten geprägt.

Weitere Anforderungen werden impliziert durch Rahmenbedingungen, die durch öffentlichen Richtlinien und Gesetze (z.B. „Reuse of Public Sector Information“ und „Informationsweiterverwendungsgesetzes“) gegeben sind.

Insgesamt lassen sich wesentliche generelle Anforderungen identifizieren, die nachfolgend näher beschrieben werden.

**Datenaktualität:** Die Aktualität der Daten ist immens wichtig für den praktischen Nutzwert der auf den Daten operierenden Anwendungen und letztendlich für die Akzeptanz dieser Anwendungen durch den Nutzer. Hier sind z.B. Verkehrsdaten und ÖPNV-Verbindungsdaten zu nennen, die einer hohen Änderungsfrequenz unterliegen. Es sind somit technische und organisatorische Vorkehrungen zu treffen, die eine hohe Datenaktualität gewährleisten. In diesem Zusammenhang kann die Definition und Überwachung von Dienstgütevereinbarungen mit Datenlieferanten sinnvoll sein. Für

den Fall, dass nicht direkt auf die Ursprungsdatenquellen zugegriffen wird, sondern auf Spiegelungen bzw. Auszüge dieser, sind Mechanismen zu implementieren, die Aktualisierungen der Originaldaten effizient erkennen und diese in entsprechende Aktualisierungen der gespiegelten Daten überführen.

**Datenverfügbarkeit:** Eine weitere Anforderung in diesem Zusammenhang ist die hohe Verfügbarkeit der Daten und ein hohes Maß an Elastizität der City Data Cloud bezogen auf das Gesamtdatenvolumen und das Volumen und die Frequenz der Datenabfragen. Zugriffszeiten und Weiterbearbeitungszeiten, z.B. mittels von der City Data Cloud bereitgestellter Integrations- und Filtermechanismen, sollten möglichst vom Datenvolumen entkoppelt werden. Entsprechende Skalierungsmechanismen sind Grundbestandteil des Cloud-Konzepts. Hinsichtlich der Datenverfügbarkeit, jedoch auch hinsichtlich der Datenaktualität, sind aber nicht nur technische Rahmenbedingungen sondern auch die durch entsprechende Richtlinien und Gesetze (PSI, Informationsweiterverwendungsgesetz) vorgegebenen Fristen und preislichen Grenzen für die Bereitstellung von Daten zu beachten.

**Datenaggregation und einheitlicher Datenzugriff:** Ein entscheidender Mehrwert entsteht häufig erst durch die Aggregation isoliert verfügbarer (Roh)Daten. Die Aggregation wird in erster Linie bedarfsbezogen durch die auf der Cloud aufsetzenden Applikationen gesteuert. Um Unternehmen und Verwaltungen bei der Entwicklung dieser Applikationen von komplexen Integrationsaufgaben zu entlasten und diesbezüglich hohe Performanz zu gewährleisten, sollte die City Data Cloud bereits Basisdienste für die technische und semantische Datenintegration anbieten. In technischer Hinsicht sollten derlei Basisdienste unabhängig von den Quellformaten die Bereitstellung von Daten in weit verbreiteten und standardkonformen Datenformaten anbieten. Zur Sicherung der semantischen Interoperabilität sollten Dienste bereitgestellt werden, die eine einheitliche Interpretation unterschiedlich repräsentierter und strukturierter Informationen ermöglichen. Die Bereitstellung der technischen Basisdienste durch die City Data Cloud-Plattform ermöglicht darüber hinaus auch ein Geschäftsfeld für Dienstleister. Diese können durch Nutzung der Basisdienste anwendungsbezogene Integrationsaufgaben als Dienstleistung anbieten und somit einen Beitrag zur Wertschöpfungskette der City Data Cloud leisten.

**Datenanalyse:** Ähnlich wie für die Datenintegration ist die gezielte Abfrage und Auswertung von Daten zu ermöglichen. Entsprechende Anforderungen an Basisdienste der City Data Cloud reichen von einfachen Abfragen über komplexere und effiziente Suchdienste bis hin zu Analysediensten für die Erkennung von Relationen, Korrelationen und Abhängigkeiten zwischen Daten. Für die Bewertung von Entwicklungen und die Analyse von Trends sind dabei auch die Verknüpfung von historischen mit aktuellen Daten zu ermöglichen. Gleichzeitig sind jedoch auch Vorkehrungen zu treffen, die eine un-erlaubte systematische Verknüpfung und Analyse von Daten unterbinden, um z.B. Prinzipien wie das Recht auf informationelle Selbstbestimmung zu berücksichtigen (siehe dazu auch „Datensicherheit“).

**Endgeräteunabhängiger Zugang zu Daten:** Unter dem Aspekt einer steigenden Endgeräteheterogenität und der wachsenden Bedeutung des ubiquitären und mobilen Zugangs zu Informationen durch Bürger und Unternehmen sollte ein weitestgehend geräteunabhängiger Zugang zu Daten ermöglicht werden. Da der Zugang aus Sicht des Nutzers in erster Linie über die Applikationen erfolgt, ist der geräteunabhängige Zugang nicht primär Aufgabe der City Data Cloud selbst. Dennoch sollten - ähnlich zu den im vorherigen Abschnitt erwähnten Integrationsfunktionalitäten - Filtermechanismen in Form von Basisdiensten bereitgestellt werden,

die z.B. eine gezielte Auswahl von Teilen einer Datenmenge bzw. eine Reduzierung der Datendichte ermöglichen.

**Datensicherheit:** Grundsätzlich stehen Offenheit und Informationsfreiheit im politischen Spannungsfeld zu Werten des Persönlichkeits- und Datenschutzes. Dieses Spannungsfeld wird insbesondere bei der Vision der City Data Cloud ersichtlich, die einerseits den Zugriff auf Daten zur gemeinsamen Gestaltung der städtischen Prozesse und Abläufe umfasst, damit jedoch auch besondere Anforderungen an die Datensicherheit stellt. Zu diesem Zweck sind unterschiedliche Schutzkategorien von Daten vorzusehen sowie geeignete Identity-Management-Mechanismen (Authentisierung, Autorisierung, Verschlüsselung), die eine effektive Steuerung der Datennutzung ermöglichen. Zudem ist die mögliche Manipulation von Daten auszuschließen. Insbesondere für Unternehmen, die auf der Basis der City Data Cloud neue Informationsprodukte und -dienste entwerfen und anbieten, ist die Integrität der Daten eine kritische Geschäftsgrundlage.

**Rechtssichere Verwendung und Weitergabe von Daten:** Für die Nutzung und Weitergabe der Daten sind transparente rechtliche Regelungen zu entwerfen und anzuwenden. Konkret bedeutet dies, dass ähnlich wie für Software auch für Daten geeignete Lizenzmodelle zu entwerfen sind. Dabei ist darauf zu achten, dass nicht ein Wildwuchs von rechtlichen Regelungen entsteht, der Unternehmen, Verwaltungen und Bürger hinsichtlich der Nutzung der Daten verunsichert bzw. künstlich einschränkt. Vielmehr sollten Standard-Lizenzen entworfen und angewendet werden, die typische Anwendungsfälle abdecken. Dabei sind auch Regelungen hinsichtlich der Verwendung und Weitergabe von höherwertigen Daten zu entwerfen, die aus der Integration/Verknüpfung/Filterung von vorhandenen Daten neu geschaffen wurden. Ansatzpunkte für derartige Lizenzwerke sind bei den Standardlizenzen der gemeinnützigen Creative Commons Gesellschaft [4] und im Bereich der Open Source Software zu finden.

**Infrastruktur für die Bereitstellung von Applikationen:** Mehrwertdienste und Applikationen, die auf der City Data Cloud aufbauen sollten einer breiten Masse von Nutzern zur Verfügung stehen. Insofern ist eine marktplatzähnliche Infrastruktur anzudenken, die das Auffinden, den Bezug und die Nutzung von Applikationen unterstützt. Es ist zu diskutieren, inwiefern die Realisierung einer solchen Infrastruktur Teil eines Umsetzungsprojekts für die City Data Cloud ist oder davon losgelöst betrachtet werden sollte. In jedem Fall sollten die Applikationen bestimmten Qualitätskriterien hinsichtlich Verfügbarkeit, Benutzbarkeit und Sicherheit genügen. Beiträge zur Sicherung der Qualität können von der einfachen Publikation von Richtlinien bis hin zur Zertifizierung von Applikationen reichen - ähnlich den Mechanismen wie sie heute von einschlägigen Marktplätzen für mobile Anwendungen bekannt sind.

### 3.1 Grobkonzeption einer City Data Cloud für Berlin

Um die Ziele der City Data Cloud Berlin umzusetzen, bedarf es einer Plattform im technischen und organisatorischen Sinne. Zum einen wird eine technische Infrastruktur zur Integration und Bereitstellung der heterogenen und verteilten Datenquellen benötigt. Zum anderen muss ein organisatorisches Rahmenwerk etabliert werden, das die Bereitstellung, Aufbereitung, Weitergabe und Nutzung der Daten für die beteiligten Akteure (Anbieter, Intermediäre und Nutzer) in rechtlicher und wirtschaftlicher Hinsicht regelt.

Im Wesentlichen kann eine City Data Cloud mit einem Vier-Schichten-Modell beschrieben werden: Die Nutzerschicht der City Data Cloud umfasst Bürger, Unternehmen und öffentliche Einrichtungen.

Diese agieren zum einen als Kunden oder Konsumenten der City Data Cloud als auch im interaktiven Sinne als Bereitsteller oder Lieferanten der Daten. Sie agieren als so genannte Prosumenten [5], denen zugleich die Rolle des Produzenten und des Konsumenten inne wohnt. Der Zugang zur City Data Cloud erfolgt dabei über eine Anwendungsschicht, in der öffentliche und kommerzielle Applikationen nutzungsspezifisch zur Verfügung stehen. Der Fokus wird hierbei auf die Innovationskraft des Marktes und der Gesellschaft gesetzt, die unterstützt durch die öffentliche Hand Mehrwertdienste und Applikation auf Basis des Datenbestandes der City Data Cloud hervorbringt. Die somit vorrangig extern bereitgestellten Applikationen kommunizieren über eine Diensteschicht mit der Smart-Communication-Plattform der City Data Cloud. Deren Kern bildet eine Infrastrukturschicht, welche die Integration und Bereitstellung von umfassenden Stadt-relevanten Daten ermöglicht. Das folgende Schaubild veranschaulicht diesen generellen Ansatz einer City Data Cloud.

Den Kern einer City Data Cloud bilden öffentliche Daten, zu denen auch Daten stadtnaher Unternehmen zur Ver- und Entsorgung gezählt werden können. Darüber hinaus bietet der Ansatz zudem die Möglichkeit, kommerzielle Daten (wie aus dem Gesundheits- und Pflegebereich oder von Handwerkern) und Daten privaten Ursprung bereitzustellen, diese miteinander Mehrwert-orientiert zu verknüpfen und über verschiedene Anwendungskanäle nutzbar zu machen.

Die technische Basisinfrastruktur einer City Data Cloud besteht wie in Abbildung 12 dargestellt aus fünf wesentlichen Basiskomponenten.

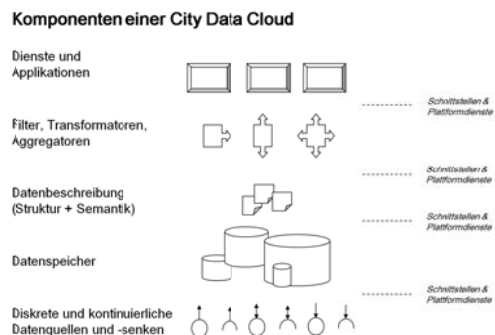


Abbildung 2 Hauptkomponenten einer City Data Cloud

Um diese Grobarchitektur anschaulich zu beschreiben, wird der Weg der Daten von der Übergabe an die City Data Cloud bis zu dessen Nutzung über entsprechende Applikationen beschrieben. Die technischen Eckpfeiler für eine Realisierung werden dabei jeweils kurz umrissen:

**Diskrete und kontinuierliche Datenquellen- und Senken:** Ein wichtiger Ansatz für den Erfolg einer City Data Cloud ist es, die bestehenden Systeme weder technisch noch organisatorisch zu beeinträchtigen. D.h. es müssen möglichst flexibel - angepasst an den jeweiligen Ursprung der Daten - verschiedene Mechanismen bereitgestellt werden, um die Daten für die City Data Cloud nutzbar zu machen. Das umfasst zum einen die Möglichkeit, Daten komplett als Kopie oder auch semi-strukturiert in Form von Spreadsheets) zu beziehen, um diese dann innerhalb der City Data Cloud weiter aufzubereiten und zu verarbeiten (diskrete Datenquelle). Zum anderen muss ebenfalls die Möglichkeit gegeben sein, an eine Organisation gebunden Daten in ihren bisherigen Systemen zu belassen und nur einen Zugriff auf die Daten durch die City Data Cloud bereitzustellen



(kontinuierliche Datenquellen). Um dies zu ermöglichen müssen wohl definierte Schnittstellen und Plattformdienste geschaffen werden, die möglichst auf Web Service Standards beruhen und einem Service-orientierten Architekturansatz folgen, um eine hohe Interoperabilität zu gewährleisten. Die Schnittstellen und Plattformdienste sollten dabei Echtzeit-basierte Infrastrukturen unterstützen, um ebenfalls sensorische Daten kontinuierlich in die City Data Cloud einspeisen zu können. Neben den Datenquellen können analog auch Datensinken bedient werden, die kontinuierlich oder diskret einen direkten Zugriff auf die Daten benötigen.

**Datenspeicher und Datenbeschreibung:** Die Daten liegen nach der Übergabe oder durch Zugriff zunächst noch in ihrer Rohfassung vor. In den Datenspeichern der City Data Cloud lassen sie sich dann möglichst automatisiert aufbereiten und strukturieren, so dass deren weitere Verarbeitungsfähigkeit erhöht wird. In diesem Kontext kann auch die explizite Datenpflege zur Verbesserung der Datenqualität als weiterer Wertschöpfungsschritt gesehen werden. Damit die Daten flexibel von möglichst vielen potentiellen Applikationen verarbeitet werden können sollten Aspekte wie: Nutzung offener Standards, Modularisierbarkeit und Skalierbarkeit sichergestellt sein [6]. Um diese Ziele zu gewährleisten, empfiehlt es sich daher dem Linked-Data-Ansatz [7] zu folgen, der vorsieht Daten in einer ausdrucksstarken Graph-Struktur zu beschreiben. Der Ansatz basiert auf offenen Datenbeschreibungsstandards des World Wide Web Konsortiums (W3C) wie RDF [8] und URI [9] der Internet Engineering Task Force (IETF). Somit ist sichergestellt, dass sämtliche heterogene Datenformate der Quellen auf beliebige Datenformate der Applikation abbildet werden können, wobei das generische RDF-Modell als flexibler Mittler gesehen werden kann. So genannte Graph-Datenbanken oder RDF-Triple-Stores dienen dabei als hochperformante persistente Datenspeicher [10]. Modularität wird durch die Linked-Data-Technologie insbesondere dadurch sichergestellt, dass unterschiedliche Vokabulare für die Datenbeschreibung und Strukturierung verwendet werden können, ohne dass ein umfassendes Datenschema über den gesamten Datenbestand der City Data Cloud vorab aufwendig abgestimmt und festgelegt werden muss. Dort wo notwendig lassen sich dann Interoperabilitätsmechanismen durch Abbildungen zwischen verschiedenen Vokabularen integrieren. Darüber hinaus wird die Skalierbarkeit bzw. Evolution der Datenbeschreibung durch die Fähigkeit der Linked-Data-Technologie gewährleistet, die es erlaubt ändernde oder neu erforderliche Terme und Definitionen flexibel zu ergänzen, ohne die bisherigen Datenstrukturen aufbrechen zu müssen oder neu aufzusetzen [6].

**Filter, Transformatoren, Aggregatoren:** Um die Daten möglichst flexible und effizient für die darauf aufsetzenden Dienste und Applikation verarbeitbar und auswertbar (Datenfusion) zu machen werden umfangreich konfigurierbare Filter, Transformatoren und Aggregatoren bereitgestellt. Diese basieren auf geeigneten Abfrage- bzw. Datenmanipulationssprachen wie SPARQL [11]. Damit lassen sich anhand von frei auswählbaren Eigenschaften und Mustern der Datenbeschreibungen die jeweils relevanten Datensätze aus der City Data Cloud herausfiltern. Das Rückgabeformat der Abfragen lässt sich dabei beliebig transformieren, so dass die Ergebnisse in gängigen Datenformaten bereitgestellt werden (z.B. strukturiertes XML oder auch JSON), wie sie für die Entwicklung von Web-Applikationen verwendet werden. Analog lassen sich somit auch flexible Aggregatoren konfigurieren, die unterschiedliche Datensätze, die auch mit heterogenen Vokabularen beschrieben sein können, zusammengeführt bereitstellen. Die Filter, Transformatoren und Aggregatoren können jeweils in Echtzeit über Dienste genutzt werden als auch anwendungsspezifisch vorverarbeitet (logisch aggregiert) oder als persistierte Replikate (physisch aggregiert) genutzt werden.

**Dienste und Applikationen:** Dienste und Applikation werden wie bereits oben erläutert vorrangig extern entwickelt und bereitgestellt. Diese umfassen klassische Web-basierte Anwendungen, die die bereitgestellten Daten der City Data Cloud in ihre Portale integrieren und ggf. mit weiteren Informationen anreichern und verknüpfen. Einen weiteren Schwerpunkt bilden mobile Applikationen des sich stark entwickelnden mobilen Internets als auch Dienste und Applikationen für spezifische eingebettete Systeme.

Insgesamt lässt sich sagen, dass die notwendigen technischen Infrastrukturkomponenten bereits vorhanden und weitgehend zudem als Open Source Software verfügbar sind.

**Datenschutz und Lizenzierung:** Neben den technischen Eckpfeilern sind rechtliche Aspekte wie Datenschutz und Lizenzierung erfolgskritisch für die Etablierung einer City Data Cloud in Berlin.

Bzgl. des Datenschutzes gibt es häufig grundsätzliche und wichtige Bedenken, wenn es um die Zusammenführung, Weiterverarbeitung und Weitergabe großer Datenmengen geht. In diesem Zusammenhang ist es wichtig klar herauszustellen, dass die Vision einer City Data Cloud nicht das Ziel hat, allumfassend Datenbestände öffentlich zu machen. Datenbestände, deren Weitergabe den Schutz der Privatsphäre oder des Rechts auf informationelle Selbstbestimmung beeinträchtigt (insbesondere personenbezogene Daten) eignen sich daher nicht für eine City Data Cloud. Davon zu differenzieren sind jedoch bereits oben genannte Daten privaten Ursprungs wie z.B. persönliche Bewertungen von Dienstleistungen oder Beiträge im Rahmen von Beteiligungsverfahren. Um die Sicherheit und Akzeptanz der City Data Cloud Berlin diesbezüglich frühzeitig zu gewährleisten, sollten deshalb bereits in der Konzeptionsphase Datenschutz-Experten mit eingebunden werden.

Ein weiterer wichtiger rechtlicher Aspekt betrifft die Lizenzierung. Befasst man sich in der Informationstechnologie üblicherweise mit Lizenzierung von Software, Dokumenten oder Multimedia-Inhalten, bedarf es für eine City Data Cloud eines Lizenzierungsmechanismus für Daten. Idealerweise sollten Lizenzierungsinformationen in einer maschinenlesbaren Form zur Verfügung stehen, um die Weiterverarbeitung der Daten möglichst effizient unter Einhaltung lizenzrechtlicher Regelungen zu ermöglichen. Es existieren bereits entsprechende Ansätze für die Lizenzierung von Daten aus dem Kontext von Open Data Initiativen, wie Open Data Commons [12]. Auf Basis existierender Ansätze sollte eine abgestimmte Auswahl an Lizenzierungsoptionen erstellt werden. Insbesondere sollten dabei die Heterogenität des Datenursprungs und verschiedene Nutzungsszenarien berücksichtigt werden und damit zusammenhängende Geschäftsmodelle (kommerzielle Lizenzen vs. nicht-kommerzielle Datenlizenzen). In diesem Zusammenhang ist ebenfalls die bereits oben diskutierte Public-Service-Information-Richtlinie zu beachten, nach der die Weiterverwendung von Informationen öffentlicher Stellen unter Gebührenvorbehalt steht.

#### **4. Die Berlin-/Brandenburger Umfrage zu öffentlichen Daten**

Im Zeitraum 23.09. bis 31.10.2010 wurde ein Online-Fragebogen zum Interesse an den Daten der öffentlichen Hand eingestellt. Dazu wurden 20 Kategorien von Abfallentsorgung bis Umweltdaten angeboten, von denen drei verschiedene Kategorien auszuwählen waren. Zudem wurde gefragt, ob man diese Daten regelmäßig im Netz lesen würde und ob man diese Daten (auch für andere) aufbereiten würde.

Es haben 1338 Bürgerinnen und Bürger ihre Stimme für bis zu 3 Kategorien abgegeben. Im Ranking ganz vorn steht das Thema „Stadtplanung“. Zweitplatziert sind Fragen zur „Verwaltung“, die heute bereits gut aufbereitet unter Berlin.de/verwaltungsführer zu finden sind. Ein aktuelles Projekt zum Wissensmanagement arbeitet derzeit noch bestehende Defizite auf. Dann folgen die Themen „Umweltdaten“ und „Kontrollen“.

Das Umfrageergebnis entsprechend der drei Auswahlrunden stellt Abbildung 3 dar, Abbildung 3a zeigt das kumulierte Umfrageergebnis. Von den Antwortenden würden 92% die bereitgestellten Daten regelmäßig lesen. 54% würden gar mit den bereitgestellten Daten konkret arbeiten, sie für andere aufbereiten und online zur Verfügung stellen. Im Vergleich zur Umfrage von SAS Deutschland [13][14], die im Zeitraum 14. bis 18.06.2010 1018 Bürger befragt hat. In dieser repräsentativen Umfrage befürworten 88% die Bereitstellung öffentlicher Daten, 81% sehen darin eine Basis für demokratische Teilhabe, 69% würden die Daten regelmäßig lesen und 7% würden mit diesen Daten arbeiten und sie für andere aufbereiten.

Im Ergebnis müssen wir unser Umfrageergebnis relativieren, da wir vorwiegend Antworten von netzaffinen Nutzern erhalten haben. Dennoch verstehen wir die Themenauswahl als wesentlich und werden entlang dessen ansetzen, den Einstieg in Open Data in Berlin zu finden.

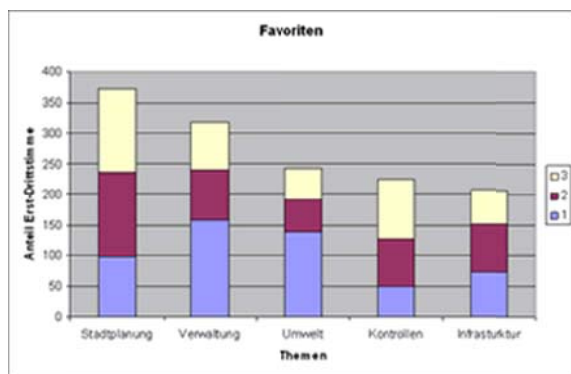


Abbildung 3 Umfrageergebnis entsprechend der drei Auswahlrunden

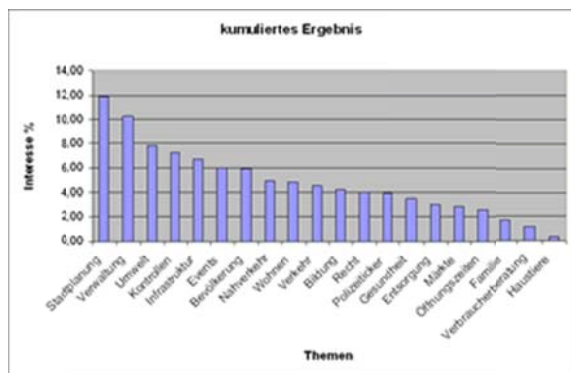


Abbildung 3a Kumuliertes Umfrageergebnis

## 5. Zusammenfassung

Im Ergebnis der Vorstudie und der Umfrage, wurden Handlungsempfehlungen zur Konzeption einer Infrastruktur für Berlin abgeleitet:

- Die Vision, in der Stadt offene Daten für die Berliner Wirtschaft und Bürger anzubieten
- Die Erarbeitung von Demonstratoren zu Fragen der Stadtplanung und/oder Umwelt im Rahmen eines Piloten
- Die Erarbeitung eines umfassenden Konzeptpapiers, das als Grundlage für eine nachhaltige Umsetzung der Berlin City Cloud dienen kann
- Die Etablierung der Berlin City Cloud als ein wesentliches Instrument zur Generierung neuer Unternehmenslösungen und neuer Unternehmen in der Stadt, welche einen wichtigen Beitrag zur wissensbasierten Re-Industrialisierung Berlins leisten kann.
- Die Etablierung der Berlin City Cloud als ein Aushängeschild für die Stadt, worauf aufbauende innovative Prozesse in Berlin ermöglicht werden und Berlin eine führende Rolle bei der Umsetzung von Open Data und Smart Cities Szenarien einnehmen kann.

Auch wenn die Konzepte einer City Cloud derzeit für und im Kontext der Stadt Berlin diskutiert werden, so sind diese Konzepte auf andere Städte übertragbar. Das europäische Projekt Open Cities [15] wurde im November 2010 mit den Städten Amsterdam, Barcelona, Berlin, Helsinki und Paris gestartet, bei dem u.a. die eine gemeinsame Konzeption für offene städtische Daten erarbeitet werden wird. Diese Arbeiten werden durch Fraunhofer FOKUS geleitet.

Zudem soll in Berlin ein Konzeptpapier den technischen, organisatorischen, regulatorischen und rechtlichen Rahmen einer City Cloud inklusive entsprechender Umsetzungsrichtlinien erarbeiten. Darüber hinaus wird ein Pilotprojekt erste offene Daten für Berlin und darauf aufbauende Szenarien prototypisch umsetzen und demonstrieren.

## Literatur und Referenzen

- [1] Open Data Principles: <http://wiki.opengovdata.org/index.php?title=OpenDataPrinciples> (zuletzt geprüft am 31.10.2010)
- [2] Open Knowledge Foundation: <http://www.okfn.org> (zuletzt geprüft am 31.10.2010)
- [3] Dr. Martin Fornefeld et al., MICUS Management Consulting GmbH (2009). Studie zur Weiterverwendung von Daten des öffentlichen Sektors (PSI) im Bereich der geografischen, meteorologischen und juristischen Informationen
- [4] Creative Commons: <http://de.creativecommons.org> (zuletzt geprüft am 31.10.2010)
- [5] Leitl, M. (2008). Was ist ein Prosument? Harvard Business Manager, Heft 3/2008, Schwerpunkt Personal, [www.harvardbusinessmanager.de/heft/artikel/a-618967.html](http://www.harvardbusinessmanager.de/heft/artikel/a-618967.html) (zuletzt geprüft am 31.10.2010)
- [6] Berners-Lee, T. (2009). Putting Government Data online - Design Issues. World Wide Web Consortium. Abgerufen Juli 28, 2010, von [www.w3.org/DesignIssues/GovData](http://www.w3.org/DesignIssues/GovData) (zuletzt geprüft am 31.10.2010)

- [7] Heath, T. (2010). Linked Data – Connect Distributed Data across the Web, [www.linkeddata.org](http://www.linkeddata.org) (zuletzt geprüft am 31.10.2010)
- [8] World Wide Web Consortium (2004). Resource Description Framework (RDF), [www.w3.org/RDF/](http://www.w3.org/RDF/)
- [9] Berners-Lee, T. R. Fielding and L. Masinter (1998). RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax IETF, [www.isi.edu/in-notes/rfc2396.txt](http://www.isi.edu/in-notes/rfc2396.txt) (zuletzt geprüft am 31.10.2010)
- [10] Rodriguez, M. A. (2009). A Reflection on the Structure and Process of the Web of Data. Bulletin of the American Society for Information Science and Technology, 35(6).
- [11] World Wide Web Consortium (2008). SPARQL Query Language for RDF. W3C Recommendation, [www.w3.org/TR/rdf-sparql-query/](http://www.w3.org/TR/rdf-sparql-query/) (zuletzt geprüft am 31.10.2010)
- [12] Jordan Hatcher (2008). Open Data Commons. Legal Solutions for Open Data, [www.opendatacommons.org](http://www.opendatacommons.org) (zuletzt geprüft am 31.10.2010)
- [13] Ina Schieferdecker, Nils Barnickel, Matthias Flügge, Steven Hager, Wolfgang Both: Vorstudie zur City Data Cloud Berlin, Fraunhofer FOKUS, Aug. 2010, [http://www.fokus.fraunhofer.de/de/motion/ueber\\_motion/arbeits Themen/smart\\_cities/downloadformular\\_cdc\\_studie/index.html](http://www.fokus.fraunhofer.de/de/motion/ueber_motion/arbeits Themen/smart_cities/downloadformular_cdc_studie/index.html) (zuletzt geprüft am 31.10.2010)
- [14] FORA Studie im Auftrag der SAS Deutschland: Open Data – Open Government MONITOR2010. Wünschen Bürger mehr Transparenz?, Aug. 2010, <http://www.sas.de/open> (zuletzt geprüft am 08.11.2010)
- [15] EU Projekt Open Cities: Projektbeschreibung, [http://ec.europa.eu/information\\_society/activities/livinglabs/docs/open\\_cities\\_pub.pdf](http://ec.europa.eu/information_society/activities/livinglabs/docs/open_cities_pub.pdf) (zuletzt geprüft am 08.11.2010)

## Autoren

### Nils Barnickel

Dipl. Inform. Nils Barnickel, Fraunhofer Institut für Offene Kommunikationssysteme, eGovernment Competence Center, Kaiserin-Augusta-Allee 31, 10589 Berlin, Tel: +49 30 3463 7131, Email: [nilsbarnickel@fokus.fraunhofer.de](mailto:nilsbarnickel@fokus.fraunhofer.de), URL: [www.fokus.fraunhofer.de/go/elan](http://www.fokus.fraunhofer.de/go/elan)

### Wolfgang Both

Dr. Wolfgang Both, Senatsverwaltung für Wirtschaft, Technologie und Frauen, Martin-Luther-Str. 105, 10825 Berlin, Tel: 030-9013-8257, Email: [wolfgang.both@senwtf.berlin.de](mailto:wolfgang.both@senwtf.berlin.de), URL: [www.projektzukunft.berlin.de](http://www.projektzukunft.berlin.de)

### Matthias Flügge

Dr.-Ing. Matthias Flügge, Fraunhofer Institut für Offene Kommunikationssysteme (FOKUS),  
eGovernment Competence Center, Kaiserin-Augusta-Allee 31, Tel. +49 30 3463 7135, Email:  
[matthias.fluegge@fokus.fraunhofer.de](mailto:matthias.fluegge@fokus.fraunhofer.de), URL: [www.fokus.fraunhofer.de/go/elan](http://www.fokus.fraunhofer.de/go/elan)

**Ina Schieferdecker**

Prof. Dr.-Ing. Ina Schieferdecker, Fraunhofer Institut für Offene Kommunikationssysteme (FOKUS),  
Competence Center Modellieren und Testen, Kaiserin-Augusta-Allee 31, Tel. +49 30 3463 7241,  
Email: [ina.schieferdecker@fokus.fraunhofer.de](mailto:ina.schieferdecker@fokus.fraunhofer.de), URL: [www.fokus.fraunhofer.de/go/motion](http://www.fokus.fraunhofer.de/go/motion)

Field Code Changed

Formatted: English (Ireland)