

# Open Data: Versunkene Schätze oder digitaler Datenmüll?

Sirko Hunnius | Bernhard Jäger

abstract

Open Data ist weltweit ein emergentes Phänomen. Dabei verlieren Nutzer oftmals den Überblick, welche Daten es wo überhaupt gibt. Das von der Europäischen Kommission geförderte FP7-Projekt OpenDataMonitor entwickelt derzeit eine Web-Plattform, welche darüber Aufschluss geben wird, wo qualitativ hochwertige offene Daten auffindbar sind, wodurch sich diese auszeichnen und wie offene Daten für eine bessere Nutzung harmonisiert werden können. Der Artikel beschreibt die methodische Vorgehensweise im Projekt und stellt erste Analyseergebnisse dar.

In den letzten Jahren hat das Thema "Open Data" zunehmend Aufmerksamkeit auf sich gezogen <sup>(1)</sup>. Immer mehr Kommunen, Länder und Staaten begannen, Teile ihrer Datenbestände zu veröffentlichen <sup>(2)</sup>. Dadurch ist allerdings eine zunehmende Unübersichtlichkeit entstanden, welche Daten wo überhaupt verfügbar sind <sup>(3)</sup>. Nationale oder europäische Portale sind nur sehr eingeschränkt in der Lage, diese Vielfalt zu konsolidieren <sup>(4)</sup>. Bisherige Ansätze, Open Data zu ordnen und zu analysieren, wie der Open Data-Index <sup>(5)</sup> oder das Open Data-Barometer <sup>(6)</sup>, sind zeitpunktbezogene Betrachtungen, die der dynamischen Entwicklung von Open Data kaum gerecht werden. Hinzu kommt, dass diese Erhebungen vielfach auf Expertenschätzungen beruhen, deren Validität unklar ist.

Im Rahmen des von der Europäischen Kommission geförderten OpenDataMonitor(ODM)-Projektes entwickelt ein Konsortium um das österreichische Technologieunternehmen SYNYO GmbH, das Potsdamer Institute for eGovernment (IfG.CC) und die Landeshauptstadt München eine Plattform, die eine umfassende Übersicht über bestehende Open Data Ressourcen liefert und zugleich Daten aus existierenden Katalogen analysiert und visualisiert <sup>(7)</sup>. Über Metadaten, Parameter und Indikatoren werden Open Data-Ressourcen differenziert analysiert und bewertet. Durch

das ODM Projekt wird so eine dynamische Analyse und Beobachtung von Open Data ermöglicht, mit automatisierten Methoden zum Absuchen von Datenkatalogen, zur Analyse von Metadaten und zur intuitiven Darstellung sowie dem Vergleich von offenen Daten.

Das Monitoring von Open Data liefert sowohl Nutzern offener Daten als auch Politik und Verwaltung selbst wertvolle Informationen: So erfahren Nutzer von Open Data, in welchen Datenportalen besonders hochwertige Datensätze bereitgestellt werden, in welchem Umfang Daten insgesamt nutzbar sind und inwieweit vergleichbare Datensätze europaweit vorliegen. Politik und Verwaltung liefert ein solches Monitoring Hinweise darauf, welche Datensätze andere Verwaltungen bereitstellen und in welcher Form diese Datensätze bereitgestellt werden (Formate, Metadaten, Datenstrukturen, Lizenzen uvam.). Der Open Data-Monitor schafft so mehr Transparenz über offene Daten und kann als Voraussetzung für eine Harmonisierung von Open Data in Europa dienen.

Die methodische Herangehensweise an das Monitoring von Open Data knüpft an den „semantics web approach“ an <sup>(8)</sup>. Demnach werden Ressourcen im Internet nach vordefinierten Schemata beschrieben, damit Maschinen in

der Lage sind, diese zu lesen und zu verstehen<sup>(9)</sup>. Diese Beschreibungen werden als Metadaten bezeichnet. Sie geben beispielsweise Auskunft darüber, wer wann einen Datensatz erzeugt hat, auf welches geografische Gebiet sich die Daten beziehen, wie häufig und wann zuletzt ein Datensatz aktualisiert wurde usw. Der Open Data-Monitor sammelt all diese Metadaten, harmonisiert diese und kann so Aussagen über Umfang, Aktualität, Abdeckung und Qualität der verfügbaren offenen Daten treffen. Damit über eine solche enorme Datenmenge (über 200 Datenportale mit jeweils mehreren tausend Datensätzen die jeweils mit einer Vielzahl von Merkmalen beschrieben sind) Aussagen getroffen werden können, wurden Metriken entwickelt, die beispielsweise die Zeitdimension von Datensätzen analysieren<sup>(10)</sup>. Hierfür wurden im Rahmen von Interviews Sichtweisen und Interessen der an Open Data beteiligten Akteure erhoben und analysiert<sup>(11)</sup>. Anschließend wurden diese Erkenntnisse durch eine großzahlige quantitative Befragung validiert. Ende Januar 2015 wurde ein Demonstrator online gestellt, der erste Analysen ermöglicht.

Erste Ergebnisse zeigen, dass die Qualität der Metadaten höchst unterschiedlich ist. So werden die Metadaten nicht nur variantenreich beschrieben; auch die Qualität der Eingaben erweckt häufig den Eindruck von Sorglosigkeit. Dies hängt mit Open Data-Policies zusammen, die unterschiedlich konkret und verbindlich sind<sup>(4)</sup>, jedoch auch mit der organisatorischen Gestaltung und wie die Verantwortlichkeiten für Open Data zugewiesen werden. Hierbei wurden häufig die Interessen derjenigen, die offene Daten erzeugen wenig beachtet<sup>(12)</sup>. Diesen entsteht oftmals tatsächlich ein hoher Aufwand (siehe auch<sup>(13)</sup>,<sup>(14)</sup>), weil selten konkrete Standards die manuelle Arbeit informieren oder gar „open by design“-Systeme die manuelle Arbeit ersetzen. Hier zeigt sich, dass Open Data Ressourcenaufwand verursacht und Aufmerksamkeit absorbiert und kein Selbstläufer ist, was in euphorischen Glaubensbekenntnissen häufig vernachlässigt wird. Aktuell richtet sich die Aufmerksamkeit der Verantwortungsträger maßgeblich darauf, ein eigenes Datenportal mit möglichst vielen Datensätzen aufzubauen. Bei der Auswahl der Datensätze dominiert zumeist der „Verfügbarkeits-Ansatz“<sup>(4)</sup>, wonach online gestellt wird, was gerade da ist: Möglichst mit geringem Aufwand, schlecht beschrieben und wenig politisch sensitiv.

Dies führt letztlich dazu, dass die vielversprechenden Datensätze zwischen dem digitalen Datenmüll kaum

auffindbar sind und verschollen zu bleiben drohen. Für die Legitimation von Open Data kann dies nach der anfänglichen Euphorie und dem Aktivismus zu einem ernsthaften Problem werden, weil so niemand die verfügbaren offenen Daten nutzen kann. Das OpenData-Monitor-Projekt soll dabei helfen, hierfür ein stärkeres Bewusstsein zu schaffen, Wege zur Harmonisierung aufzuzeigen und den ein oder anderen Datensatz zu Tage fördern. ■



**Sirko HUNNIUS,**  
Dipl.-Kfm. (FH)  
Wissenschaftlicher  
Mitarbeiter, Institute for  
eGovernment (IfG.CC);  
shunnius@ifg.cc



**Mag. Bernhard JÄGER**  
Research Manager/Open-  
DataMonitor Projektkoordi-  
nator, SYNYO GmbH;  
bernhard.jaeger@synyo.  
com

## literatur & links

<sup>(1)</sup> Meijer, A. J., de Hoog, J., Van Twist, M., van der Steen, M., & Scherpenisse, J. (2014). Understanding the Dynamics of Open Data: From Sweeping Statements to Complex Contextual Interactions. In M. Gascó-Hernandez (Hrsg.), *Open Government. Opportunities and Challenges for Public Governance* (S. 101-114). New York: Springer.

<sup>(2)</sup> Heimstädt, M., Saunderson, F., & Heath, T. (2014). From Toddler to Teen: Growth of an Open Data Ecosystem. *eJournal of eDemocracy & Open Government JeDEM*, 6 (2), 123-135.

<sup>(3)</sup> Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 28 (4), 258-268.

<sup>(4)</sup> Hunnius, S., Krieger, B., & Schuppan, T. (2014, September). Providing, Guarding, Shielding: Open Government Data in Spain and Germany. In 2014 EGPA Annual Conference, Speyer, Deutschland.

<sup>(5)</sup> Open Knowledge Foundation. *Open Data Index*. 2013. (2014, Juli). Verfügbar unter: <https://index.okfn.org/country>

<sup>(6)</sup> Davies, T. (2013) *Open Data Barometer*. 2013 Global Report, London.

<sup>(7)</sup> OpenDataMonitor. (2015, Jänner). Verfügbar unter: <http://project.opendatamonitor.eu>

<sup>(8)</sup> Berners-Lee, T., & Hendler, J. (2001). Publishing on the semantic web. *Nature*, 410, 1023-1025.

<sup>(9)</sup> Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The Semantic Web Revisited. *Intelligent Systems IEEE*, 21(3), 96-101.

<sup>(10)</sup> Atz, U. (2014). The Tau of Data: A New Metric to Assess the Timeliness of Data in Catalogues. *Proceedings of the International Conference for E-Democracy and Open Government (CeDEM2014)*, 2014.

<sup>(11)</sup> OpenDataMonitor Project. (2014). *D2.4 Open Data Stakeholder Requirement Report 1*, Vienna.

<sup>(12)</sup> Hunnius, S., & Krieger, B. (2014, August). The Social Shaping of Open Data through Administrative Processes. *Proceedings of the 10th International Symposium on Open Collaboration OpenSym*, Berlin, Deutschland.

<sup>(13)</sup> Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, 19, 129-152.

<sup>(14)</sup> Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical Impediments of Open Data. *Electronic Journal of e-Government*, 10 (2), 156-172.