

OPEN DATA ANALYTICS AS A SERVICE

Klaus-Peter Eckert, Matthias Flügge, Stephan Gauch



Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

1. Auflage April 2014

Coverfoto: © mirpic/ Fotolia.com

Alle Rechte vorbehalten

© Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS

Fraunhofer-Institut für Offene
Kommunikationssysteme FOKUS
Kaiserin-Augusta-Allee 31
10589 Berlin

Telefon: +49-30-3436-7115

Telefax: +49-30-3436-8000

elankontakt@fokus.fraunhofer.de

www.fokus.fraunhofer.de

Dieses Werk ist einschließlich aller seiner Teile urheberrechtlich geschützt. Jede Verwertung, die über die engen Grenzen des Urheberrechtsgesetzes hinausgeht, ist ohne schriftliche Zustimmung des Instituts unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen sowie die Speicherung in elektronischen Systemen. Die Wiedergabe von Warenbezeichnungen und Handelsnamen in diesem Buch berechtigt nicht zu der Annahme, dass solche Bezeichnungen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und deshalb von jedermann benutzt werden dürften.

Soweit in diesem Werk direkt oder indirekt auf Gesetze, Vorschriften oder Richtlinien (z.B. DIN, VDI) Bezug genommen oder aus ihnen zitiert worden ist, kann das Institut keine Gewähr für Richtigkeit, Vollständigkeit oder Aktualität übernehmen.

ISBN

978-3-00-046007-4

OPEN DATA ANALYTICS AS A SERVICE

Klaus-Peter Eckert, Matthias Flügge, Stephan Gauch

April 2014

Fraunhofer-Institut für Offene Kommunikationssysteme

Der wissenschaftlichen Partner Fraunhofer FOKUS bedankt sich herzlich beim ISPRAT e.V. für die Unterstützung zur Erstellung der vorliegenden Studie.

Inhaltsverzeichnis

INHALTSVERZEICHNIS	I
ABBILDUNGSVERZEICHNIS	III
ABKÜRZUNGSVERZEICHNIS	IV
EXECUTIVE SUMMARY	V
1 EINLEITUNG	1
1.1 Ausgangssituation sowie ein kurzer Überblick zum Stand des Wissens	1
1.2 Problemstellung und Ziel der Studie	2
1.3 Einführung	2
2 GRUNDPRINZIPIEN VON OPEN DATA	4
2.1 Open Government Data	6
2.2 Rechtliche Grundlagen	8
2.3 Praxisbeispiele	10
2.4 Technologische Aspekte	11
3 GRUNDPRINZIPIEN VON CLOUD-ARCHITEKTUREN	13
3.1 Cloud-Charakteristiken	13
3.2 Cloud-Dienstmodelle	14
3.3 Cloud Bereitstellungsmodelle	16
3.4 NIST Cloud Service Taxonomie	17
3.5 NIST Cloud-Definition	18
3.6 BSI Cloud-Definition	18
3.7 NIST Akteure	19
3.8 NIST Anwendungsfälle	20
3.9 NIST Referenzarchitektur	22
4 BESTANDTEILE VON ODAAAS	25
4.1 Datenermittlung - Retrieve	27
4.2 Datenbereinigung - Revise	29
4.3 Datenzusammenführung - Integrate	32
4.4 Datenanalyse - Analyze	35
4.5 Akteure im Umfeld von ODAaaS	38
5 NUTZEN VON DATENANALYSE IM BEREICH OPEN DATA	40
5.1 Übergeordnete Nutzenpotentiale	40
5.2 Kompetenzen der ODAaaS-Akteure	42
5.3 Umsetzungskonzepte für ODAaaS	46
6 EXPERTENANFORDERUNGEN AN ODAAAS	49
6.1 Methodische Vorbemerkungen	49

6.2	Ergebnisse der Experteninterviews	49
6.2.1	Metadaten	50
6.2.2	Barrierefreiheit	50
6.2.3	Verfügbarkeit und Aktualität offener Daten	50
6.2.4	Datenqualität	51
6.2.5	Hohes Potential durch Big Data und Sensordaten	52
6.2.6	Die Rolle einer Community bei der Analyse offener Daten	53
6.2.7	Open Data Analytics – einfache vs. komplexe Konzepte	53
6.2.8	Abrechnungsmodelle für ODaaS	54
6.2.9	Apps für die Analyse offener Daten	54
6.2.10	Spezifische Anforderungen für ODaaS	55
7	VARIANTEN DER ANALYSE OFFENER DATEN IN DER CLOUD	56
8	HANDLUNGSEMPFEHLUNGEN	59
8.1	Maßnahmen bei der Bereitstellung offener Daten	59
8.2	Maßnahmen bei der Analyse offener Daten	60
8.3	Empfehlungen für Anbieter von AaaS-Werkzeugen	61
9	ANHANG – FRAGEBOGEN	63
9.1	Technische Aspekte	63
9.1.1	Nutzerverwaltung und Barrierefreiheit	63
9.1.2	Datenhaltung und -verarbeitung	63
9.2	Datenschutz und Kontrollfunktionen.	64
9.3	Betrieb und Geschäftsmodelle	65
9.4	Allgemeine Fragestellungen	65
10	LITERATURVERZEICHNIS	67

Abbildungsverzeichnis

Abbildung 1: Beziehungen zwischen den Begrifflichkeiten im Cloud Computing	17
Abbildung 2: NIST Cloud-Dienste Taxonomie	18
Abbildung 3: NIST Cloud-Referenzarchitektur ²³	22
Abbildung 4: Prozessschritte und zugehörige Anforderungen in ODAaaS	26
Abbildung 5: Kompetenzen der Akteure in ODAaaS	42
Abbildung 6: (O)DAaaS in der Cloud-Referenzarchitektur	57
Abbildung 7: Datenanalyse durch kombinierte Cloud-Dienste	58

Abkürzungsverzeichnis

AaaS	Analytics as a Service
AGPL	Affero General Public License
BI	Business Intelligence
BSI	Bundesamts für Sicherheit in der Informationstechnologie
CCUCDG	Cloud Computing Use Case Discussion Group
CKAN	Comprehensive Knowledge Archive Network
CSA	Cloud Security Alliance
CSV	Comma-separated values
DCAT	Data Catalog Vocabulary
DMTF	Distributed Management Task Force
eGovG	E-Government-Gesetz
ETL	Extraction-Transform-Load
ELT	Extraction-Load-Transform
FDBS	Federated Database System
IaaS	Infrastructure as a Service
IFG	Informationsfreiheitsgesetz
IKT	Informations und Kommunikationstechnologie
IWG	Informationsweiterverwendungsgesetz
NIST	National Institute of Standards and Technology
ODA	Open Data Analytics
ODAAaaS	Open Data Analytics as a Service
OKF	Open Knowledge Foundation
PaaS	Platform as a Service
PSI	Public Sector Information
RDF	Resource Description Framework
SaaS	Software as a Service
SAJACC	Standards Acceleration to Jumpstart Adoption of Cloud Computing
SOA	Service Oriented Architecture
SPARQL	SPARQL Protocol And RDF Query Language
URL	Uniform Resource Locator
VM	Virtuelle Maschine
XML	Extensible Markup Language

Executive Summary

Im Zeitalter der Informationsgesellschaft sind Daten und Informationen immer leichter, schneller und in größeren Mengen verfügbar. Wissensbasierte Dienstleistungen bilden eine der wesentlichen Grundlagen für das Wirtschaftswachstum der Industrienationen der vergangenen 20 Jahre. Mit der digitalen Bereitstellung und dem Einsetzen semantischer Technologien werden Informationen zwar immer schneller bearbeitbar, gleichzeitig werden in den Teilbereichen moderner Gesellschaften immer mehr Daten produziert. Dies gilt auch für den Bereich offener Daten, d.h. für Daten, die von jedermann frei verwendet, nachgenutzt und verbreitet werden können.

Daten der öffentlichen Verwaltung sowie aus Wissenschaft und Forschung werden sowohl national als auch international zunehmend als **Open (Government) Data** bereitgestellt. Die Bereitstellung dieser Daten soll einerseits durch Transparenz das Vertrauen in das staatliche Handeln verstärken und Missbrauch oder Fehlentscheidungen einschränken. Andererseits soll durch die Entwicklung von diese Daten analysierenden **Data Analytics**-Anwendungen ein Mehrwert für die Gesellschaft generiert werden. Insbesondere für Verwaltungen sind hohe Effizienzgewinne möglich, sofern strategische Entscheidungen unter Einbeziehung möglichst automatisierter Auswertungen offener Daten getroffen werden. Verwaltungsübergreifender Informationsaustausch, aktuelle Bürgerdienste und die Bereitstellung von Informationen an Unternehmen können bei Bedarf auf Knopfdruck erfolgen.

Ziel der vorliegenden Studie ist es, die Potentiale und Hindernisse bei Entwicklung, Bereitstellung und Nutzung von Data Analytics-Anwendungen **as a Service** als Cloud-Dienste aufzeigen. Für existierende Anwendungen wird deren Eignung zum Umgang mit den Formaten und Schnittstellen analysiert, über die typischerweise Daten auf Open Data-Plattformen bereitgestellt werden. Es wird untersucht, inwieweit sich die Anwendungen in Open Data-Plattformen integrieren lassen.

Die Studie gibt einführend einen Überblick über den aktuellen Stand der Technik für Open Data und Cloud Computing. Ausgehend von typischen Prozessschritten bei der Analyse offener Daten werden die Nutzenpotentiale für die beteiligten Akteure ermittelt und Architekturkonzepte für die Umsetzung von **Open Data Analytics as a Service** in lokalen und föderierten Cloud-Infrastrukturen diskutiert. Die erarbeiteten Ergebnisse werden an Expertenmeinungen von Anbietern von Analytics-Anwendungen gespiegelt und in Handlungsempfehlungen für die Bereitstellung und Analyse offener Daten sowie Entwicklung zugehöriger Werkzeuge zusammengefasst.

1 Einleitung

1.1 Ausgangssituation sowie ein kurzer Überblick zum Stand des Wissens

Im Zeitalter der Informationsgesellschaft sind Daten und Informationen immer leichter, schneller und in größeren Mengen verfügbar. Wissensbasierte Dienstleistungen bilden eine der wesentlichen Grundlagen für das Wirtschaftswachstum der Industrienationen der vergangenen 20 Jahre. Mit der digitalen Bereitstellung und dem Einsetzen semantischer Technologien werden diese Informationen zwar immer schneller bearbeitbar, gleichzeitig werden in den Teilbereichen moderner Gesellschaften immer mehr Daten produziert: Die Wissenschaft erstellt z.B. in der Genomanalyse oder der molekularen Teilchenforschung Daten in zuvor unbekanntem Ausmaß. In der Wirtschaft werden im Zuge der Herstellung intelligenter Produkte immer größere Datenmengen erzeugt und verarbeitet. Im öffentlichen Sektor erzeugen beispielsweise Sensornetzwerke zur Messung und Beobachtung von Umweltzuständen laufend neue Datensätze.

Existierende Trends zur Bereitstellung dieser Daten zur Untersuchung und Verwendung durch Dritte machen diese Datenmengen auch immer leichter für neue Vorhaben verfügbar. So bieten bereits verschiedene Initiativen des öffentlichen Sektors in entsprechenden Webportalen Datensätze in maschinenlesbaren Formaten und unter Lizenzen an, welche eine freie Weiterverwendung erlauben. Von speziellem Interesse sind hier die sogenannten „Open Data“, d.h. Datenbestände, die im Interesse der Allgemeinheit der Gesellschaft ohne jedwede Einschränkung zur freien Nutzung, zur Weiterverbreitung und zur freien Weiterverwendung frei zugänglich gemacht werden.

Gerade kleine und mittelständische Unternehmen, zivilgesellschaftliche Organisationen oder einzelne Verwaltungseinheiten sehen sich vor der Anforderung, diese Daten zu für ihre eigenen Zwecke sinnvoll zu analysierenden und individuell auszuwertenden Informationen aufzubereiten sowie auf offenen Datenbeständen aufsetzende, neue innovative Apps/Mash-ups für Bürger zu bauen.

Am Markt für Datenanalyse (englisch: data analytics) existieren zunehmend dedizierte Lösungen zur Speicherung, Veröffentlichung und Aufbereitung von hochvolumigen „Big Data“, also von schnell wachsenden Datenmengen, die sowohl strukturiert in Form von Zahlen und Tabellen wie auch unstrukturiert in Form von Texten, Videos und Websites vorliegen. Technologien aus Bereichen wie z.B.

- Speicherung dynamischer, einfach strukturierter Massendaten,
- „Linked Open Data“ oder „Topic Maps“ zur Beschreibung offener Daten und deren semantischen Zusammenhängen,
- „Business Intelligence“ bzw. „Data Analytics“ zur systematischen Datenanalyse,
- „Cloud Computing“ zur Bereitstellung leistungsfähiger IKT-Infrastrukturen

stellen einzelne Bausteine auf dem Weg zu einer holistischen Aufbereitung und Bereitstellung derartiger Daten und aus diesen abgeleiteten Informationen dar. Innovative, auf offenen Daten aufbauende Mehrwert-Anwendungen werden erst durch die Analyse und Verschneidung von Daten aus unterschiedlichen Quellen möglich. Ein umfassender Ansatz, der unter Berücksichtigung technischer, betrieblicher und rechtlicher Rahmenbedingungen den Zugriff und die Bewertung der Daten und aus diesen abgeleiteter Analysen ermöglicht, ist jedoch in vielen Fällen noch nicht möglich.

Die Studie untersucht die Anforderungen aber auch Möglichkeiten für „Open Data Analytics as a Service“ und vergleicht heute bereits vorhandene Ansätze auf ihre Eignung, die Auswertung offener Daten als Software-Dienst zu ermöglichen.

1.2 Problemstellung und Ziel der Studie

Die primäre in der Studie behandelte Fragestellung lautet:

Wie kann bei stetig steigenden Datenmengen die Auswertung offener Daten für verwaltungsinterne und -externe Akteure einfach gestaltet werden?

Ein Ziel dieser Untersuchungen ist es, Potentiale und Hindernisse für Werkzeuge zur Datenanalyse als Dienstleistung aufzeigen. Für existierende Data Analytics-Werkzeuge wird deren Eignung zum Umgang mit den Formaten und Schnittstellen analysiert, über die typischerweise Daten auf Open Data-Plattformen bereitgestellt werden. Weiterhin wird untersucht, inwieweit sich die Werkzeuge in existierende Plattformen integrieren lassen.

Insbesondere für Verwaltungen sind hohe Effizienzgewinne möglich, sofern Entscheidungen auf Basis weitestgehend automatisierter Auswertungen offener Daten getroffen werden. Verwaltungsübergreifender Informationsaustausch, aktuelle Bürgerdienste oder die Bereitstellung von Informationen an Unternehmen können bei Bedarf auf Knopfdruck erfolgen. Informationsgewinn auf Knopfdruck birgt jedoch auch spezielle Risiken, auf die in der Studie detailliert eingegangen wird. Datenanalyse als Dienst kann nur funktionieren, sofern die eingesetzten Werkzeuge kurze Reaktionszeiten besitzen. Dies stellt spezielle, in der Studie zu detaillierende, Anforderungen an die genutzte IKT-Infrastruktur.

Die schnelle Bereitstellung von Informationen darf nicht darüber hinwegtäuschen, dass das Analyseergebnis Entscheidungsprozesse erleichtert, nicht jedoch vorausnimmt. Der Charakter eines „Business Support Systems“ darf auch bei „Open Data Analytics - ODA“ nicht verloren gehen. Entsprechende Anforderungen werden in der Studie herausgearbeitet und mögliche Anwendungsfälle für die Analyse offener Daten beschrieben.

1.3 Einführung

Der Austausch von Wissen und Informationen hat für unsere Gesellschaft immer mehr an Bedeutung gewonnen. Die dezentrale Verteilung und Bereitstellung dieser wichtigen Ressourcen werden als Zugpferd für die Weiterentwicklung demokratischer und freier Gesellschaften angesehen. Die Open Data-Bewegung greift diesen Umstand auf und plädiert seit mehreren Jahren für eine kostenfreie Bereitstellung und Wiederverwendung von wissenschaftlich und gesellschaftlich nutzbaren Daten.

Die Offenlegung von Daten ist jedoch nur ein erster Schritt, um ihre Potentiale vollständig zu nutzen. Bereits heute werden auf Basis von offenen Daten Apps entwickelt, welche den Nutzen der in Open Data verborgenen Informationen für ausgewählte Fragestellungen deutlich zeigen. Dies ist jedoch aus Sicht der heutigen Möglichkeiten, Daten nutzbringend einzusetzen, nur ein erster Schritt. Durch Verwendung von Analysewerkzeugen kann das Potential offener Daten wesentlich breiter erschlossen werden.

Im Rahmen dieser Studie soll aufgezeigt werden, welche Anforderungen im Rahmen der Umsetzung von „Open Data Analytics as a Service“ (ODAAaaS) möglich sein können und welche Potentiale und Anforderungen sich

daraus ergeben. Besondere Berücksichtigung gilt dabei dem Einsatz von Cloud-Technologien als technischer Infrastruktur.

Im Rahmen der Studie wurden Experteninterviews mit Anbietern aus dem Bereich Datenanalyse/Analytics durchgeführt und daraus technische, organisatorische und wirtschaftliche Anforderungen für die Etablierung von ODAAaaS identifiziert. Aus den Anforderungen wurden zugehörige Handlungsempfehlungen abgeleitet.

Die Studie gibt in den Kapiteln 2 und 3 Einführungen in die Grundprinzipien offener Daten und des Cloud Computing. Kapitel 4 entwickelt ein Prozessmodell aus vier Schritten zur Beschreibung der wichtigsten Phasen bei der Analyse offener Daten. Dieses Prozessmodell dient im Kapitel 5 zur Diskussion der Potentiale von ODAAaaS für die beteiligten Akteure.

Kapitel 6 diskutiert das vorgeschlagene Prozessmodell und dessen Potentiale aus Sicht der befragten Anbieter von Analysewerkzeugen. In Kapitel 7 werden über heutige Angebote hinausgehende, technische Lösungsvorschläge für die Analyse offener Daten unter Nutzung von Cloud-Technologie vorgestellt. Aus den durchgeführten Interviews und den vorgeschlagenen Lösungsansätzen abgeleitete Handlungsempfehlungen schließen die Studie in Kapitel 8 ab.

2 Grundprinzipien von Open Data

Daten der öffentlichen Verwaltung werden sowohl national als auch international in zunehmender Weise als Open Data bereitgestellt. Dabei wird angeführt, dass die Bereitstellung von Daten aus der öffentlichen Verwaltung sowie aus Wissenschaft und Forschung einerseits durch Transparenz das Vertrauen in das staatliche Handeln verstärken und Missbrauch oder Fehlentscheidungen einschränken kann und andererseits durch die Entwicklung von auf diesen Daten basierenden Anwendungen ein großer Mehrwert für die Gesellschaft generiert wird.

Auch die Europäische Kommission schätzt das Potenzial von Open Data auf mehreren Ebenen hoch ein. So heißt es in ihrer Open Data-Strategie¹: *„Diese Informationen haben ein beträchtliches und derzeit ungenutztes Potenzial für die Weiterverwendung in neuen Produkten und Dienstleistungen und für Effizienzsteigerungen in den Verwaltungen. Aus der Öffnung dieser Ressource könnte sich ein gesamtwirtschaftlicher Nutzen von bis zu 40 Mrd. EUR jährlich in der EU ergeben. Die Öffnung öffentlicher Datenbestände wird auch eine stärkere Beteiligung der Bürger am politischen und gesellschaftlichen Leben ermöglichen und bestimmten Politikbereichen (z.B. Umwelt) zugutekommen.“*

Damit diese Potential ausgeschöpft werden kann, legt die Europäische Kommission Maßnahmen für die Öffnung von Informationen des öffentlichen Sektors (englisch: Public Sector Information, PSI) fest. Diese reichen von der Anpassung des bestehenden Rechtsrahmens über die Mobilisierung von Finanzinstrumenten bis zur Koordinierung des Erfahrungsaustausches zwischen den Mitgliedsstaaten.

Nach einer Definition der Open Knowledge Foundation (OKF) sind Offene Daten *„[...] Daten, die von jedermann frei verwendet, nachgenutzt und verbreitet werden können – maximal eingeschränkt durch Pflichten zur Quellennennung und ‘share-alike’ (Weitergabe unter gleichen Bedingungen).“*² Danach muss ein Datensatz mehrere Bedingungen aufweisen, um als offen angesehen werden zu können.

Insbesondere haben sich dabei die maschinelle Verarbeitbarkeit, die freie Lizenzierung und der einfache Zugang zu den Daten als ausschlaggebende Merkmale bei der effektiven Weiterverwendung von offenen Daten herausgestellt.³

Maschinelle Verarbeitbarkeit

Die Verwendung von maschinenlesbaren Formaten bei der Veröffentlichung von frei verfügbaren Daten ist ein wesentliches Kriterium, um solche Informationen tatsächlich als offen einstufen zu können. Erst die Maschinenlesbarkeit vereinfacht die Aufbereitung der Daten für unterschiedliche Anwendungen. Lieben z.B. tabellarische Daten in PDF-Dokumenten vor, so können zwar Menschen diese Daten problemlos lesen und verstehen, da dieses Format auf die Qualität der Bildschirmanzeige ausgelegt ist, aber die maschinelle Lesbarkeit und Weiterverwendbarkeit ist nur eingeschränkt gegeben. Im Gegensatz dazu stellen allgemein von Software interpretierbare Formate wie XML (Extensible Markup Language), CSV (Comma Separated Values) oder RDF (Resource Description Framework) sicher, dass die Daten von einer großen Anwendergruppe effizient weiterverwendet und aufbereitet werden können.

Bei der Wahl der einzusetzenden Formate ist zusätzlich zu der Wiederverwendbarkeit, der Zugang zu diesen Datenstrukturen zu berücksichtigen. Hierfür dürfen plattformspezifische Bedingungen oder andere formattypische Vorgaben nicht die Weiterverwendung eingrenzen, da innovative, auf offenen Daten aufbauende Werkzeuge und

¹ Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss DER Regionen – Offene Daten: Ein Motor für Innovation, Wachstum und transparente Verwaltung (KOM/2011/0882)

² <http://www.opendefinition.org>

³ Für eine ausführliche Darstellung der Voraussetzungen für Offene Daten siehe <http://opendefinition.org/okd/deutsch/>

Anwendungen erst durch die Analyse und Verschneidung von Daten aus unterschiedlichen Quellen ermöglicht werden. Folglich sollten für einen uneingeschränkten Einsatz der publizierten Inhalte verbreitete offene Standards verwendet werden, die vollständig dokumentiert und öffentlich zugänglich sind.

Freie Lizenzierung

Bei der Kennzeichnung von Informationen als „offen“ im Sinne von Open Data ist diese Bezeichnung mit der uneingeschränkten und kostenfreien Weiterverwendbarkeit der Inhalte assoziiert. Trotz der bisher unternommenen Versuche diesen Begriff zu definieren, gibt es bislang keine einheitliche Erklärung zum Verständnis der Offenheit von Daten.

Aus den Open Data-Leitlinien lässt sich allerdings schlussfolgern, dass die freie Lizenzierung der Beiträge ähnlich wie bei der maschinellen Verarbeitbarkeit einen entschiedenen Einfluss auf die Verarbeitung der Daten hat. Erhebt beispielsweise die für die Veröffentlichung der Informationen zuständige Stelle Geldleistungen für die weitere Verwendung der Daten, so werden damit potentielle Nutzergruppen ausgeschlossen.

Außerdem können auch die in den Nutzungsbedingungen aufgeführten Restriktionen eine Vielzahl von Dateninteressenten an der Weiternutzung der verfügbaren Informationen hindern.

Hindernisse bei der Wiederverwendung von offenen Datensätzen können sich auch auf bestimmte Zielgruppen oder einzelne Arbeitsvorgänge beziehen. Eine Untersagung der kommerziellen Nutzung der Daten oder der Aufbereitung der Ursprungsdaten entspricht nicht einer angestrebten, freien Lizenzierung. Dementsprechend kann jede Einschränkung in Bezug auf die Weiterverwendung der Daten durch Dritte den Kreis der Nutzer erheblich eingrenzen und sollte weitestgehend vermieden werden.

Da grundsätzlich Daten unter Urheberrechten stehen, die die Nutzung und Verarbeitung durch Dritte regulieren, gilt es bei der Veröffentlichung von Informationen eine geeignete Lizenz sorgfältig zu bestimmen. Im Rahmen der Open Data-Bewegung hat sich eine Reihe von Lizenzen herauskristallisiert, die in diesem Zusammenhang hilfreich sein können. Die wichtigsten in diesem Zusammenhang zu nennenden Lizenzen sind die Creative Commons – Namensnennung⁴, die Open Data Commons Namensnennung für Daten⁵ und die im Rahmen des deutschen nationalen Datenportals entwickelte Datenlizenz Deutschland für Daten deutscher Verwaltungen⁶.

Einfacher Zugang

Zwar existieren bislang sehr viele öffentlich zugängliche Daten, die in analoger Form wie beispielsweise als Papierdokumente oder als digitale Datenbanken vorliegen, jedoch sind diese oftmals in geschlossenen Strukturen verfügbar, d.h. nur wenige Personen können diese Daten praktisch einsehen und weiterverarbeiten. Obwohl diese Informationen freigegeben sind und keine rechtlichen Beschränkungen für die Nutzung bestehen, wird dennoch das effiziente Finden dieser Daten nicht unterstützt, da das Datenmaterial nur lokal an den jeweiligen Verwaltungsgrenzen abrufbar ist. Insofern ist der Zugang zu den Daten nur über eine Anfrage an die zuständige Stelle möglich.

Daher ist es ein Anliegen des Open Data-Ansatzes, den Prozess der Informationsbeschaffung schneller und einfacher zu gestalten. Insbesondere zielt die Veröffentlichung der Daten darauf ab, die beim Zugang anfallende Hürden technischer, organisatorischer sowie rechtlicher Art wesentlich zu reduzieren. Um den damit verbundenen Ansprüchen nach einem einheitlichen und leichten Zugriff auf offene Daten gerecht zu werden, werden die relevanten Datenbestände über zentral zugängliche Webportale auffindbar gemacht. Um ein effizientes Auffinden

⁴ <http://www.creativecommons.org/licenses/cc-by/3.0>

⁵ <http://www.opendefinition.org/licenses/odc-by>

⁶ http://www.daten-deutschland.de/bibliothek/Datenlizenz_Deutschland/dl-de-by-1.0

zu ermöglichen, werden die Datensätze mit Metadaten, wie Titel, Quelle und Nutzungsbedingungen, ausgezeichnet.

Der stark vereinfachte Zugang zu offenen Informationen ist im Hinblick auf die zielgerechte Verwendung der veröffentlichten Daten ein essentielles Kriterium. Sie bildet zusammen mit der maschinellen Verarbeitbarkeit und der freien Lizenzierung die Grundvoraussetzungen für die künftige Entwicklung neuer Anwendungen, die auf offenen Daten basieren.

2.1 Open Government Data

Mit dem rasanten Fortschritt von Informations- und Kommunikationstechnologien steigen auch die Erwartungen der Bürger, sowie von Wissenschaft und Wirtschaft, an Qualität, Zugänglichkeit und Offenheit von Leistungen der öffentlichen Hand. Die Umsetzung dieser Erwartungen setzt einen grundlegenden Wandel im Verwaltungshandeln und -denken voraus. Dem wurde von Seiten der Bundesregierung mit der Aufnahme von Open Government in das Regierungsprogramm „Vernetzte und transparente Verwaltung“ Rechnung getragen.

Open Government bezeichnet dabei vielfältige Öffnungsprozesse von Staat und Verwaltung mit der Maßgabe Transparenz, Partizipation und Kollaboration zwischen Staat, Bürgern, Wissenschaft und Wirtschaft zu fördern und somit einen öffentlichen Diskurs über die Effektivität und Effizienz staatlichen Handelns zu führen.

Um einen stärkeren Diskurs und darauf aufbauende Interaktion zwischen den Akteuren zu ermöglichen, ist neben transparenteren Prozessen und Entscheidungen eine Bereitstellung von Informationen und Daten, die dem Staat zur Verfügung stehen und fortwährend im großen Maße erhoben werden, wesentlich.

Diese Öffnung der Verwaltungsdaten wird als Open Government Data bezeichnet und kann wie folgt definiert werden: *„Offene Verwaltungsdaten sind jene Datenbestände des öffentlichen Sektors, die von Staat und Verwaltung im Interesse der Allgemeinheit ohne jedwede Einschränkung zur freien Nutzung, zur Weiterverbreitung und zur freien Weiterverwendung frei zugänglich gemacht werden“*.⁷

Mit der Offenlegung von Daten des öffentlichen Sektors sind drei wesentliche Ziele verbunden, die in Teilen auch für offene Daten des Privatsektors gelten: Transparenz, Innovation und Effizienz.

Transparenz

Unter Transparenz wird im Allgemeinen die Möglichkeit verstanden, *„außerhalb von Einrichtungen Einsicht in Prozesse und Entscheidungen zu nehmen, um sich selbst eine Meinung dazu bilden zu können“*.

Bislang blieben viele Daten, sofern sie nicht über allgemein zugänglichen Quellen auffindbar waren, nur ausgewählten Kreisen vorbehalten. Oftmals wurde der Zugang zu solchen Informationen für Dritte lediglich auf eine Anfrage hin und nach Einschätzung der zuständigen Stellen herausgegeben.

Damit allerdings der erwünschte Meinungsbildungsprozess zu Stande kommen kann, müssen sich Außenstehende angemessen informieren können, was über eine aktive Offenlegung von relevanten Informationen und Daten erreicht wird. Eine hierbei oft gewählte Herangehensweise umfasst zur Etablierung dauerhafter transparenter Strukturen die Bereitstellung von diesen Datenbeständen über geeignete webbasierte Datenportale.

Mit den damit gewonnen Informationen erhalten Bürger die Möglichkeit, bestimmte Vorgänge und Entscheidungen zu verstehen und nachzuvollziehen, wodurch ihre informierte Entscheidungsfindung gefördert

⁷ von Lucke und Geiger (2010)

wird. Darüber hinaus trägt die transparente Darlegung von frei verfügbaren Daten dazu bei, bestehende Strukturen effektiv zu verbessern, indem Ideen, Wünsche und Rückmeldungen von Außenstehenden zu den Daten in die Optimierungsprozesse eingebunden werden. Neben der verstärkten Interaktion und Zusammenarbeit fördert Transparenz außerdem die Außendarstellung von den verantwortlichen Stellen, die durch die verstärkte Sichtbarkeit, welches mit der Veröffentlichung von neuen Inhalten einhergeht, erzielt wird.

Innovation

Durch das Verfügbarmachen von frei zugänglichen Daten ergeben sich für verschiedene Nutzergruppen unterschiedliche Chancen diese Beiträge produktiv zu verwerten. Mit offenen (Verwaltungs-)Daten kann neues Wissen generiert und diese als Grundlage für kreative Innovationen eingesetzt werden.

Dabei kann der Einsatzzweck sehr vielfältig angelegt sein. Unter anderem kann die Verarbeitung geeigneter Datensätze einerseits im öffentlichen Bereich zur Bewältigung gesellschaftlicher Anforderungen und andererseits mit einem wirtschaftlichen Hintergrund zur Entwicklung kommerzieller Applikationen dienen. In diesem Zusammenhang entfalten sich auch für die Wissenschaft neue Perspektiven und Nutzungsmöglichkeiten das neue Datenangebot sinnvoll in laufende und künftige Forschungstätigkeiten zu integrieren.

Unabhängig davon in welcher Domäne offene Inhalte bei der Implementierung von innovativen Lösungen Anwendung finden, profitieren alle Zielgruppen von den Vorzügen frei verfügbarer Daten. Beispielsweise tragen Lösungsansätze für die Verbesserung existierender Infrastrukturen zum Gemeinwohl der Gesellschaft bei, wohingegen Unternehmen, die ihre Geschäftsmodelle auf frei zugänglichen Daten aufsetzen, neue Arbeitsplätze schaffen und das Wachstum der Wirtschaft fördern. Darüber hinausgehend steht es für alle Bürgerinnen und Bürger frei mit dem verfügbaren Datenmaterial eigenständig neue Anwendungen und Dienste für ihre individuellen Fragestellungen und Bedürfnisse zu entwickeln.

In den vergangenen Jahren wurde bereits anhand zahlreicher Applikationen demonstriert, in wie fern offene Datensätze Mehrwert schaffen können. Dennoch ist das Potenzial frei zugänglicher Informationen bislang nicht vollständig ausgeschöpft, da täglich die Menge an Datenquellen steigt und den Weg für neue und bisher unentdeckte Einsatzgebiete bereitet.

Effizienz

Eine als Konsequenz von der Veröffentlichung von frei verfügbaren Daten resultierende Effizienzsteigerung ist unmittelbar in Verwaltungsprozessen zu beobachten. Mit dem vereinfachten Zugriff auf offene Informationen über öffentlich zugängliche Datenportale ist eine Reduzierung entsprechende Nachfragen an Verwaltungsmitarbeiter zu erwarten. Somit erhalten Fachkräfte die Möglichkeit, mehr Zeit für Kernaufgaben aufzuwenden.

Zum anderen kann mit der Offenlegung von Daten der Arbeitsaufwand interner Abläufe und Austauschprozesse reduziert werden. In großen Organisationen ist häufig nur lückenhaft bekannt, welche Informationen in den verschiedenen Organisationseinheiten erhoben werden bzw. verfügbar sind. Open Data-Prinzipien und Infrastrukturen wirken diesem Problem entgegen, da für Abteilung A ersichtlich wird, über welche Daten Abteilung B verfügt.

Weiterhin eröffnen Datenportale, die frei zugängliche Informationen bereitstellen, die Möglichkeit die Verifizierung dieser Ressourcen voranzutreiben und daher die Qualität der Daten insgesamt zu verbessern. Durch den öffentlichen Zugang können Dritte Fehler und Unstimmigkeiten in den Daten aufdecken und eine entsprechende Korrektur veranlassen.

Die Qualitätssicherung, die bislang von einem kleinen Kreis von Personen durchgeführt wurde, kann folglich durch die Offenlegung der Daten erheblich verbessert werden.

Der größte erhoffte Produktivitätsgewinn, der mit der Öffnung von Daten in Verbindung steht, liegt im Potential des durch die Daten generierten Wissens begründet. Auf Basis von Rückmeldungen und Verbesserungsvorschläge seitens der Datennutzer können Prozesse und Entscheidungsvorgänge verbessert und beschleunigt werden.

2.2 Rechtliche Grundlagen

Die Vorgaben zur Weiterverwendung von Informationen des öffentlichen Sektors in Deutschland sind in der PSI-Richtlinie (2013/37/EU), dem Informationsweiterverwendungsgesetz (IWG) und dem E-Government-Gesetz (EgovG) geregelt.

Europäische Richtlinien

Mit der Veröffentlichung der Richtlinie 2013/37/EU wurde ein weiterer Schritt in Richtung der Öffnung europäischer Verwaltungsdaten unternommen. Die Richtlinie ist eine Überarbeitung der 2003 veröffentlichten Richtlinie zur Weiterverwendung von Informationen des öffentlichen Sektors⁸, die bereits wesentliche Punkte zur Förderung Offener Daten beinhaltet.

Hauptziel dieser Richtlinien ist die Schaffung von Bedingungen zur Förderung der Entwicklung unionsweiter Dienstleistungen basierend auf Daten der öffentlichen Hand. Diese Produkte und Dienstleistungen müssen den Wettbewerbsvorschriften der Europäischen Union entsprechen. Derzeit herrschen aufgrund unterschiedlicher nationaler Bestimmungen und Verfahren der Mitgliedsstaaten bei der Bereitstellung und Weiterverwendung öffentlicher Daten rechtliche Unklarheiten. Die Richtlinien stellen somit ein Mindestmaß an Regeln dar, um zu verhindern, dass diese Unsicherheiten hemmend auf die Entstehung von grenzüberschreitenden Produkten und Dienstleistungen einwirken.

Die Mitgliedsstaaten werden mit dieser Richtlinie verpflichtet, Verwaltungsdaten für die kommerzielle und nicht-kommerzielle Weiterverwendung zur Verfügung zu stellen. Diese sollen zudem unter Bedingungen bereitgestellt werden, die die Weiterverwendung so wenig wie möglich einschränken. So sollen offene Lizenzen gefördert werden, die eine Weiterverwendung ohne technische, finanzielle oder geografische Einschränkungen zulassen.

Diese Vorgaben zur Offenheit und Transparenz für öffentliche Stellen und ihrer Arbeitsweise sollen von den Mitgliedsstaaten binnen zwei Jahren in ihr nationales Recht übernommen werden.

Zwar unterstützt die Richtlinie die nicht-diskriminierende Weiterverwendung von Regierungsdaten und fordert Rechte der Bürgerinnen und Bürger bei der Antragsstellung auf Zugang zu Dokumenten und Informationen öffentlicher Stellen ein, doch ist es den Verwaltungen weiterhin gestattet, Gebühren für die Bereitstellung zu erheben, um ihre Kosten zu decken. Diese Gebühren dürfen in der Regel maximal ihre Ausgaben für die Vervielfältigung, das Anbieten und die Verbreitung der Informationen betragen. Somit ist ein wesentlicher Punkt des Open Data-Ansatzes zwar noch nicht vollständig in die Europäischen Richtlinien eingeflossen, doch stellt die Novellierung eine weitere, wichtige Verbesserung und Vereinheitlichung der Rahmenbedingungen innerhalb der europäischen Staatengemeinschaft dar, um die Wertschöpfung von offenen (Verwaltungs-)Daten weiter voranzutreiben und Potenziale nutzbar zu machen.

⁸ Vgl. <http://eur-lex.europa.eu/lexuriserv/lexuriserv.do?uri=oj:l:2003:345:0090:0096:de:pdf>

Informationsfreiheitsgesetz

Der weitreichende Zugang zu Informationen ist ein wichtiger Baustein für die Legitimation und Transparenz staatlichen Handelns. Die gewünschte Partizipation ziviler Bevölkerungsgruppen an politischen Entscheidungsprozessen kann zudem nur erreicht werden, wenn die Bürgerinnen und Bürger eines Landes mit ausreichend Informationen ausgestattet sind, auf deren Grundlage sie sich einbringen und Entscheidungen mittragen können.

Durch das Gesetz zur Regelung des Zugangs zu Informationen des Bundes (Informationsfreiheitsgesetz – IFG)⁹ wurde 2006 in Deutschland auf Bundesebene diesem Gedanken Rechnung getragen. Es gewährt jedem Bürger ein Recht auf freien Zugang zu Informationen öffentlicher Stellen des Bundes. Dieser Zugang ist voraussetzungslos und muss vom Antragssteller nicht mehr durch ein berechtigtes Interesse begründet werden. Es findet somit eine Umkehr in der Denkweise statt, so dass Behörden ggf. nun berechtigte Gründe, die einer Offenlegung entgegenstehen, darlegen müssen. Zu diesen Tatbeständen gehören u.a. Schutz der inneren Sicherheit, Schutz des behördlichen Entscheidungsprozesses, Schutz von personenbezogenen Daten oder Schutz des geistigen Eigentums. Diese Ausnahmetatbestände decken sich mit den Grundsätzen von Open Data.

Allerdings sind die Behörden nach diesem Gesetz nicht verpflichtet ihre Behördenakte und Informationen proaktiv zur Verfügung zu stellen und werden nur punktuell auf Verlangen aktiv. Des Weiteren können – mit Ausnahme von kleinen Auskünften und bei Ablehnung des Antrags – Gebühren erhoben werden. Diese Gebühren sind zwar so zu wählen, dass sie auf den Bürger nicht abschreckend wirken, doch stellen sowohl der Verwaltungsakt des Antrages, die Wartezeiten und die aufkommenden Gebühren für die Zivilgesellschaft Barrieren dar, die dem Open Data-Gedanken entgegenstehen.

Aufgrund der föderalen Strukturen sind im Informationsfreiheitsgesetz des Bundes nur die Datenbestände nationaler Behörden adressiert. Um auch auf die Informationen der Landesbehörden zugreifen zu können, sind entsprechende Gesetze auf Landesebene notwendig. Derzeit haben lediglich 11 Bundesländer Informations- und Transparenzgesetze erlassen, die es dem Bürger erlauben, die Bereitstellung von Informationen ihrer öffentlichen Stellen zu beantragen. Als positives Beispiel kann hier das Bremer Informationsgesetz herausgestellt werden, das seiner öffentlichen Verwaltung durch eine Novellierung 2011 des IFG vorschreibt, „*Verwaltungsvorschriften von allgemeiner Bedeutung*“ proaktiv in einer zentralen Datenbank zu veröffentlichen. Dazu gehören u.a. Handlungsempfehlungen, Statistiken, Gutachten und Informationen, zu denen aufgrund individueller Anträge Zugang gewährt wurden. Entsprechende gesetzliche Regelungen sind auch in Hamburg verabschiedet worden.

E-Government-Gesetz

Das Gesetz zur Förderung der elektronischen Verwaltung (EGovG / E-Government-Gesetz¹⁰), trat im August 2013 in Kraft. Es soll einerseits die Kommunikation mit der Verwaltung erleichtern und andererseits Bund, Ländern und Kommunen ermöglichen, einfachere, nutzerfreundlichere und effizientere elektronische Verwaltungsdienste anzubieten. Das Gesetz strebt dabei in seiner Gänze eine umfassende Modernisierung der Arbeitsweise des öffentlichen Sektors an und ermöglicht Verwaltungen fortschrittliche Technologien in ihre Verwaltungspraxis einzuführen.

Neben der Verpflichtung zur Eröffnung elektronischer Kommunikationskanäle (De-Mail¹¹), einer Einführung der elektronischen Aktenführung und elektronischer Bezahlungsmöglichkeiten, wird explizit die Bereitstellung von maschinenlesbaren Datenbeständen durch die Verwaltung festgelegt. Durch § 12 Abs. 1 EGovG werden Behörden nun verpflichtet, Daten bei denen ein Weiterverwendungsinteresse zu erwarten ist, grundsätzlich in maschinenles-

⁹ http://www.bmi.bund.de/DE/Themen/Moderne-Verwaltung/Open-Government/Informationsfreiheitsgesetz/informationsfreiheitsgesetz_node.html

¹⁰ http://www.bmi.bund.de/DE/Themen/IT-Netzpolitik/E-Government/E-Government-Gesetz/e-government-gesetz_node.html

¹¹ http://www.cio.bund.de/Web/DE/Innovative-Vorhaben/De-Mail/de_mail_node.html

baren Formaten über ein öffentlich zugängliches Netz zur Verfügung zu stellen. Mit der Maschinenlesbarkeit wird ein inhärentes Kriterium für die Bereitstellung offener Daten erfüllt.

Auch die freie Lizenzierung, die als maßgeblich für die Öffnung von (Verwaltungs-)Daten angesehen wird, kann von der Bundesregierung durch Rechtsverordnungen mit Zustimmung des Bundesrats geregelt werden (§ 12 Abs. 3 EGovG).

Viele Bundesländer arbeiten derzeit an einer Umsetzung der Regelungen des E-Government-Gesetzes auf Länderebene.

2.3 Praxisbeispiele

Im Vergleich zu anderen Staaten steht Deutschland in Bezug auf Open Data am Anfang seiner Entwicklung. So haben u.a. die Vereinigten Staaten und Großbritannien schon frühzeitig damit begonnen, die Grundsätze von Open Data in ihr Verwaltungshandeln zu übernehmen und zentrale Open Data-Portale¹² aufzubauen, die einen unkomplizierten Zugriff auf Verwaltungsdaten ermöglichen. Die beiden Portale sind vorbildhafte Beispiele für die Umsetzung des Open Data-Gedankens. Gestützt durch den Gesetzgeber, die eine rasche Umsetzung vorsahen, wurde eine Vielzahl von Datensätzen aus den verschiedenen Ressorts der Regierungsbehörden zentral bereitgestellt.

Die Datenportale fungieren dabei nicht nur als reine Datenbanken von Verwaltungsinformationen. Zusätzlich werden Werkzeuge und passende Schnittstellen bereitgestellt, mit deren Hilfe sich Daten visualisieren oder sinnvoll miteinander verknüpfen lassen. Während durch die Veranstaltung von Ideenwettbewerben (z.B. Apps4America) die Entwicklung von Apps und Mash-Ups gefördert wird, zeigt ein „Showroom“ bereits realisierte Projekte und entstandene Anwendungen, die den gewonnenen Mehrwert Offener Verwaltungsdaten veranschaulicht. In Foren können zudem neue Ideen und Vorschläge diskutiert und Entwicklungspartner für bevorstehende Projekte gefunden werden. Daraus ergibt sich eine Community, die basierend auf Offenen Daten innovative Produkte und Dienstleistungen für die Gemeinschaft erstellt.

Diese Vorgehensweise wird auch in anderen europäischen Ländern übernommen, doch zeigt sich hier die Entwicklung weitaus weniger rasant. Das ist auf europäischer Ebene sicherlich mit der zuvor stattfindenden Harmonisierung der einzelnen nationalen Vorschriften zu begründen. Des Weiteren wurde die Thematik Open (Government) Data und deren Realisierung in den USA und Großbritannien zu einem frühen Zeitpunkt von höchster politischer Stelle unterstützt. Das verlieh der Umsetzung der Open Data-Prinzipien innerhalb der Verwaltung einen höheren Stellenwert und führte somit zu schnelleren Ergebnissen.

Auf europäischer Ebene kann das EU-Projekt Open Cities¹³ angeführt werden, das sich zum Ziel gesetzt hat, offene und anwendergetriebene Innovationsansätze für den öffentlichen Sektor umzusetzen. In einer Kooperation von Industriepartnern, Forschungsinstituten und Universitäten aus sieben europäischen Metropolen wurde auf der Basis verschiedener Werkzeuge, Studien und Komponenten, z.B. in den Bereichen Crowd Sourcing und Open Data, eine Plattform zur Bereitstellung offener Daten entwickelt. Ergebnisse der einzelnen Arbeitspakete flossen bereits in regionale Datenportale (z.B. Amsterdam und Berlin) sowie auf nationaler Ebene in das Datenportal Govdata.de ein. Open Cities fördert so den grenzüberschreitenden Erfahrungsaustausch in Bezug auf Open (Government) Data und den internationalen Transfer von Wissen und technologischen Lösungen.

¹² <http://data.gov> und <http://data.gov.uk>

¹³ <http://opencities.net>

Dass sich offene Daten nicht ausschließlich auf Regierungsdaten beschränken müssen, zeigt das Beispiel des Open Data-Portals der Weltbank.¹⁴ Die Weltbank veröffentlicht umfassende Statistiken und Indikatoren zum Entwicklungsstand von Ländern aus aller Welt. Dabei wird die Entwicklung von Anwendungen basierend auf bereitgestellten Daten ebenfalls durch Visualisierungstools unterstützt. Um das Ziel der Entwicklungshilfe und Armutsbekämpfung voranzutreiben stellt die Weltbank zusätzlich eine Vielzahl wissenschaftlicher Publikationen und Forschungsergebnisse aus ihren Forschungsprojekten zur freien Weiterverwendung zur Verfügung.

2.4 Technologische Aspekte

Nachfolgend werden typische technische Komponenten einer Open Data-Plattform beschrieben. Zudem werden Eigenschaften von Metadaten und Datenformaten im Kontext Open Data diskutiert.

Typische Komponenten einer Open Data-Plattform

Aus technischer Sicht bestehen Open Data-Plattformen i.d.R. aus den Komponenten „Datenportal“, „Datenregister“ und „Datenspeicher“.

Datenbereitsteller, wie öffentliche Stellen und Unternehmen, beschreiben die von Ihnen bereitgestellten offenen Daten mittels Metadaten (Ansprechpartner, Zeitbezug, Ortsbezug, Lizenz), die in einem Datenregister gespeichert und verwaltet werden. Die eigentlichen Datensätze werden häufig dezentral, z.B. auf entsprechenden kommunalen Webseiten, durch die Datenbereitsteller verwaltet und zugänglich gemacht. In diesen Fällen verweisen die Metadaten lediglich auf die dezentral vorgehaltenen Datensätze, z.B. mittels einer Web-Adresse (URL).

Die Inhalte des Metadatenregister werden über ein Web-basiertes Open Data-Portal dargestellt und durchsuchbar gemacht. Direkte Datennutzer verwenden das Open Data-Portal, um Datensätze zu finden und über entsprechende Verweise auf die dezentral vorgehaltenen Daten zuzugreifen. Zudem bietet ein Open Data-Portal üblicherweise Informationen zu Anwendungen, die basierend auf den offenen Daten entwickelt wurden, sowie Interaktionsmöglichkeiten, wie z.B. die Kommentierung von Datensätzen oder Diskussionsforen rund um das Thema Open Data.

Optional kann eine Open Data-Plattform auch Funktionalitäten für die zentrale Speicherung und Abfrage der eigentlichen Daten anbieten. In diesem Fall wird als Teil der Plattform ein Datenspeicher eingerichtet. Dieser ermöglicht im einfachsten Fall das Ablegen und Herunterladen ganzer Dateien oder auch weitergehende Funktionalitäten, wie die strukturierte Abfrage von Daten über entsprechende Schnittstellen. Ein solcher Datenspeicher ist besonders dann relevant, wenn es Datenbereitsteller gibt, die nicht über die Möglichkeit verfügen, Daten auf einem eigenen Webserver bereitzustellen oder wenn es sich z.B. um besonders große Datenmengen handelt.

Die heute am häufigsten verwendete (z.B. daten.berlin.de, data.gov.uk, govdata.de, open-data.europa.eu) Basissoftware für die Realisierung von Open Data-Plattformen ist das Softwareprodukt CKAN¹⁵ der Open Knowledge Foundation¹⁶. CKAN ist Open Source-Software, die unter der AGPL-Lizenz kostenfrei zur Verfügung steht und von einer aktiven Entwicklercommunity weiterentwickelt wird. CKAN bietet bereits heute zahlreiche Erweiterungen, z.B. für den Import von Geodaten, sowie für die Datenvisualisierung und Datenspeicherung. CKAN verfügt über

¹⁴ <http://data.worldbank.org>

¹⁵ CKAN, <http://ckan.org/>

¹⁶ Open Knowledge Foundation, <http://okfn.org/>

dokumentierte Web-basierte APIs für die Abfrage und Verwaltung von Metadateneinträgen. CKAN-Instanzen sind zudem förderbar.

Metadaten und Datenformate

Wie oben bereits mehrfach hervorgehoben, sind vor allem die Metadaten, die für die Beschreibung von offenen Daten verwendet werden, von zentraler Relevanz. Anhand der Metadaten kann ein Datennutzer relevante Daten finden und erhält zusätzliche Informationen, die eine Wiederverwendung und Verarbeitung der Daten erleichtern. Für die Beschreibung von offenen Daten mit Metadaten hat sich noch kein einheitlicher Standard durchgesetzt. Es empfiehlt sich jedoch eine Orientierung an Best-Practice-Beispielen, wie data.gov.uk, govdata.de oder daten.berlin.de. Im nationalen Kontext ist diesbezüglich die Metadatenstruktur der Open Government Data-Plattform govdata.de hervorzuheben. Die dort verwendete Metadatenstruktur¹⁷ wurde mit zahlreichen Akteuren auf Landesebene und kommunaler Ebene, sowie mit Betreibern von anderen Fachdatenportalen (Geodaten, statistische Daten etc.) abgestimmt.

Im Rahmen einer W3C-Arbeitsgruppe wird derzeit ein Vokabular für Datenkataloge „Data Catalog Vocabulary“ (DCAT) entwickelt. Die Spezifikation¹⁸ hat aktuell den Status einer „Candidate Recommendation“ und ist somit auf dem Weg zu einem offiziellen W3C-Standard. Aufgrund unterschiedlicher lokaler Anforderungen ist zu erwarten, dass es trotz intensiver Harmonisierungsbemühungen, auch in Zukunft immer Unterschiede in den Metadatenbeschreibungen verschiedener Open Data-Plattformen geben wird. Hier sind Abbildungen auf gemeinsame Vokabulare, wie z.B. DCAT, ein probates Mittel für die Föderation von Open Data-Portalen und für eine plattformübergreifende Suche.

Letztendlich sollten auch die technischen Formate in denen offene Daten publiziert werden, gewissen Kriterien genügen. Zwei der von der „Open Government Working Group“¹⁹ publizierten Open Government Data-Prinzipien²⁰ treffen diesbezüglich Aussagen und wurden oben bereits herausgestellt: Offene Daten sollten so strukturiert sein, dass sie automatisch verarbeitbar („maschinenbearbeitbar“) sind. Zudem sollten die Datenformate nicht proprietär sein, d. h. die Verwendung offener Standards wird vorausgesetzt. Zudem hat der WWW-Erfinder Tim Berners-Lee ein 5-Sterne-System²¹ für „Linked Open Data“ beschrieben. Datensätze, die den o. g. Prinzipien genügen, d. h. die in maschinenverarbeitbaren und nicht-proprietären Formaten vorliegen, erhalten demnach jedoch nur drei Sterne. Volle fünf Sterne gibt es laut Berners-Lee, wenn semantische W3C-Standards, wie RDF²² und SPARQL, für die Bereitstellung bzw. Abfrage verwendet werden und die Daten zudem nicht isoliert vorliegen, sondern miteinander semantisch verknüpft sind. Obwohl die technischen Vorteile von Linked Open Data offensichtlich sind, ist der Anteil der offenen Daten, die in dieser Form bereitgestellt werden in der Praxis noch vergleichsweise gering.

Zusammenfassend kann man festhalten, dass sich aus rein technischer Sicht drei Dimensionen ergeben, entlang derer Daten mehr oder weniger offen sein können: 1) Die strukturierte Abfragemöglichkeit von Daten, 2) die Verwendung offener Standards für Datenformate und 3) die semantische Ausdrucksstärke. So würden sich, um konkrete Beispiele zu nennen, auf einer Offenheitsskala gescannte Dokumente als digitale Bilder am einen Ende (wenig offen), tabellarische CSV-Daten im Mittelfeld, und RDF Daten, die über SPARQL abfragbar sind, am anderen Ende (sehr offen) befinden.

¹⁷ Govdata.de Metadatenstruktur, <https://www.govdata.de/metadaten/schema>

¹⁸ Data Catalog Vocabulary (DCAT), <http://www.w3.org/TR/vocab-dcat/>

¹⁹ Open Government Working Group, https://public.resource.org/open_government_meeting.html

²⁰ Eight Principles of Open Government Data, https://public.resource.org/8_principles.html

²¹ Linked Open Data, <http://www.w3.org/DesignIssues/LinkedData.html>

²² Resource Description Framework, <http://www.w3.org/RDF/>

3 Grundprinzipien von Cloud-Architekturen

In diesem Kapitel wird ein Überblick über den aktuellen Stand der Wissenschaft im Bereich Cloud Computing gegeben. Dabei spielen die Definitionen des amerikanischen National Institute of Standards and Technology (NIST) (NIST - National Institute of Standards and Technology, 2013) eine herausragende Rolle. Darüber hinaus sind die konzeptionellen Arbeiten des Cloud Incubators der Distributed Management Task Force (DMTF) (DMTF - Distributed Management Task Force, 2009) und der Arbeitsgruppe WG3 im ISO/IEC SC38 (Distributed application platforms and services – cloud computing) (ISO/IEC JTC1/ SC38/WG3, 2013) besonders hervorzuheben. Aus Anwendersicht ist weiterhin die Berücksichtigung der Ergebnisse der Cloud Computing Use Case Discussion Group (CCUCDG) (CCUCDG - Cloud Computing Use Case Discussion Group, 2010) wichtig. Zur Definition und Einordnung von „Analytics as a Service“ (AaaS) in die Cloud Begriffswelt ist jedoch eine Beschränkung auf die NIST-Definitionen hinreichend.

Die Definitionen des NIST sind die heutzutage allgemein akzeptierten Beschreibungen für Cloud Computing. Die Definitionen wurden im Jahr 2009 vorgestellt (Mell & Grance, The NIST Definition of Cloud Computing, 2009) und Anfang 2011 (Mell & Grance, The NIST Definition of Cloud Computing (Draft), 2011) offiziell publiziert. Sie umfassen drei wesentliche Aspekte. Cloud Charakteristiken (englisch: essential characteristics) beschreiben wesentliche Eigenschaften von Cloud Computing. Die Dienstmodelle (englisch: service model) identifizieren die heute gebräuchlichen Arten von Cloud-Diensten. Spezielle Deployment-Modelle beschreiben unterschiedliche Bereitstellungsmodelle für Cloud-Infrastrukturen. Ein guter Überblick über diese Konzepte wird unter anderem im Eckpunktepapier des Bundesamts für Sicherheit in der Informationstechnologie BSI (BSI - Bundesamt für Sicherheit in der Informationstechnik, 2011) vom Mai 2011 gegeben. Wegen ihrer großen Bedeutung sollen die Begriffe und ihr Zusammenhang jedoch auch an dieser Stelle beschrieben werden.

Der Begriff „Cloud-Infrastruktur“ ist ein zentraler Begriff für die folgenden Definitionen. Er bezeichnet die Menge an Hardware und Software, die die fünf wesentlichen Charakteristiken des Cloud Computing ermöglicht. Eine Cloud-Infrastruktur besteht aus einem physikalischen und einem Abstraktions-Layer. Der physikalische Layer umfasst diejenigen Hardware-Ressourcen, die benötigt werden, um die angebotenen Cloud-Dienste zu erbringen. Der Abstraktions-Layer umfasst die auf dem physikalischen Layer bereitgestellte Software die benötigt wird, um die Cloud-Charakteristiken zu implementieren.

3.1 Cloud-Charakteristiken

Aus Sicht des NIST sind fünf Cloud-Charakteristiken hervorzuheben (Mell & Grance, The NIST Definition of Cloud Computing (Draft), 2011)

- **On-demand self-service:** Die Bereitstellung von Ressourcen durch den Cloud-Anbieter erfolgt für deren Nutzer transparent/automatisch. (*„A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service’s provider.“*)
- **Broad network access:** Cloud-Dienste können über standardisierte Internet-Mechanismen und Protokolle genutzt werden. Sie nicht an einen bestimmten Klienten gebunden (*„Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms e.g., mobile phones, laptops, and PDAs.“*)

- **Resource pooling:** Ressourcen werden unterschiedlichen Nutzern mandantenfähig zur Verfügung gestellt. Der geographische Ort der Ressourcen bleibt dabei im Allgemeinen verborgen. Nutzer können aber vertraglich den Speicherort, also z.B. Region, Land oder Rechenzentrum, festlegen. („*The provider’s computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction e.g., country, state, or data-center. Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.*“)
- **Rapid elasticity:** Ressourcen werden dem Nutzer schnell und flexibel, in manchen Fällen auch automatisch, zur Verfügung gestellt, so dass sich im Allgemeinen der Eindruck unendlicher Kapazitäten ergibt. („*Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.*“)
- **Measured Service:** Die Nutzung von Ressourcen ist messbar. Sie kann überwacht und zur Kontrolle von Dienstleistungsvereinbarungen und zur Erstellung von Abrechnungen genutzt werden. („*Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service.*“)

Ergänzende Eigenschaften wurden 2009 durch die Cloud Security Alliance - CSA (CSA - Cloud Security Alliance, 2009) vorgeschlagen. Dabei wurde speziell Wert auf die Mandantenfähigkeit von Cloud-Diensten und die Verwendung von an SOA (Service Oriented Architecture) angelehnte Architekturen gelegt. Weiterhin werden der Lebenszyklus von Cloud-Diensten sowie diesen unterstützende Werkzeuge als wichtiges Thema angesehen.

Abrechenbarkeit von Cloud-Diensten

Die Messbarkeit der Nutzung von Ressourcen durch Cloud-Dienste in Verbindung mit deren Überwachung, Kontrolle und Abrechnung gehört zu den wesentlichen Eigenschaften von Cloud-Systemen. Beschreibungen auf welche Art und Weise die Nutzung gemessen, überwacht und abgerechnet werden kann findet man beispielsweise im *Commercial Framework* der *Open Data Center Alliance* (Open Data Center Alliance, 2013). Dort wird unter anderem ein Überblick über verschiedene Tarifmodelle im Cloud Computing gegeben.

Eine ressourcenbasierte Abrechnung von in einer Cloud-Infrastruktur bereitgestellten AaaS-Diensten ist mit den im Business Support, siehe Kapitel 3.9 und (NIST - National Institute of Standards and Technology, 2011), bereitgestellten Funktionen Pricing und Rating möglich. Proprietäre AaaS-Dienste benötigen dagegen eigene Abrechnungsmechanismen.

3.2 Cloud-Dienstmodelle

Die Cloud-Dienstmodelle beschreiben unterschiedliche Arten von Cloud-Diensten. Ein „Cloud-Dienst“ wird sehr allgemein als eine durch den Anbieter des Dienstes ermöglichte „Befähigung des Dienstnutzers“ definiert. Diese Befähigung kann nun, je nach Dienstmodell darin bestehen, beliebige Software auf einer Menge von fundamentalen Ressourcen wie Rechenleistung, Speicher oder Netzwerk bereitzustellen und auszuführen (IaaS), eigene

Anwendungen in einer Menge von vorgegebenen Sprachen, Bibliotheken, Software-Komponenten und Werkzeugen bereitzustellen und auszuführen (PaaS) oder auf der Cloud-Infrastruktur bereitgestellte Anwendungen zu nutzen (SaaS).

Je nach Art werden diese Dienste unterschiedlichen Zielgruppen wie Betreibern, Entwicklern und Endnutzern angeboten. Die drei durch NIST vorgeschlagenen Dienstmodelle (Mell & Grance, The NIST Definition of Cloud Computing (Draft), 2011) sind allgemein akzeptiert. Darüber hinaus werden beispielsweise Infrastrukturdienste weiter differenziert und beispielsweise die Bereitstellung von Speicherkapazität als „Storage as a Service“ bezeichnet. Am anderen Ende des Spektrums können Software-Dienste beispielsweise durch Prozessdienste (englisch: Business Process as a Service), Datenzugriffe (englisch: Data as a Service), Informationsbereitstellung (englisch: Information as a Service) oder auch Analysedienste (englisch: Analytics as a Service) erweitert

- **Cloud Infrastructure as a Service (IaaS):** IaaS konzentriert sich auf die Bereitstellung von IT-Ressourcen wie Rechenleistung, Speicherkapazität oder Netzwerken als Cloud-Dienste. Der Nutzer kann beliebige Software in dieser Infrastruktur bereitstellen und ausführen ohne direkten Einfluss auf das Management der Infrastruktur zu besitzen. Der typische Nutzer eines IaaS ist daher der Betreiber von in der Cloud gehosteten Anwendungen. (*„The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls)“*).
- **Cloud Platform as a Service (PaaS):** Dem Nutzer wird eine Entwicklungs- und Laufzeitplattform mit dedizierten Diensten und nichtfunktionalen Eigenschaften zur Verfügung gestellt. Der Nutzer kann die Plattform konfigurieren, besitzt jedoch keinen Zugriff auf die zugrundeliegende Infrastruktur. Der typische Nutzer eines PaaS ist daher der Entwickler und Betreiber von in der Cloud gehosteten Anwendungen. (*„The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.“*)
- **Cloud Software as a Service (SaaS):** Dem Nutzer werden über Standard Web-Schnittstellen zugreifbare Anwendungen zur Verfügung gestellt. Der typische Nutzer eines SaaS ist daher der Endnutzer von in der Cloud gehosteten Anwendungen. (*„The capability provided to the consumer is to use the provider’s applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.“*)

3.3 Cloud Bereitstellungsmodelle

Die vier Bereitstellungsmodelle (englisch: deployment models) unterscheiden, auf welche Art und Weise Cloud-Dienste bereitgestellt und betrieben werden (Mell & Grance, The NIST Definition of Cloud Computing (Draft), 2011). Sie sind insbesondere für die Definition von Geschäftsmodellen und Sicherheitsarchitekturen wichtig.

- **Private cloud:** In privaten Clouds wird die Cloud-Infrastruktur speziell für einen einzelnen Kunden betrieben. Sie kann von dem Kunden selbst oder einem Dritten organisiert und geführt werden und kann dabei im Rechenzentrum der eigenen Institution oder einer fremden Institution stehen. („*The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise*“).
- **Community cloud:** In Community Clouds wird die Cloud-Infrastruktur für eine exklusive Menge von Kunden mit gemeinsamen Zielen betrieben. Sie kann von einer oder mehreren beteiligten Institution oder einem Dritten organisiert und geführt werden und kann dabei in eigenen Rechenzentren oder in einer fremden Institution stehen. Man kann sich hier beispielsweise gemeinsame Rechenzentren für mehrere Gebietskörperschaften vorstellen. („*The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.*“)
- **Public cloud:** In der öffentlichen Cloud wird die Cloud-Infrastruktur durch einen externen Dienstleister (kommerziell, öffentlich, forschungsorientiert) für mehrere unabhängige Kunden betrieben. („*The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.*“)
- **Hybrid cloud:** Hybride Clouds beschreiben eine Kombination aus mindesten zwei der bereits erwähnten Bereitstellungsmodelle. Dabei ist die Portabilität von Daten und Anwendungen zwischen den beteiligten Cloud-Infrastrukturen gewährleistet. („*The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds.*“)

In Abbildung 1 werden die eingeführten Begriffe und deren Zusammenhang in Form einer „Mind Map“ grafisch dargestellt. Ziel der folgenden Überlegung ist es, AaaS zu diesen Begriffen in Relation zu setzen. Zuvor wird die Begriffswelt jedoch noch um Definitionen von Cloud Computing selber, typischen Akteuren und Anwendungsfällen sowie um ein architektonisches Referenzmodell ergänzt.

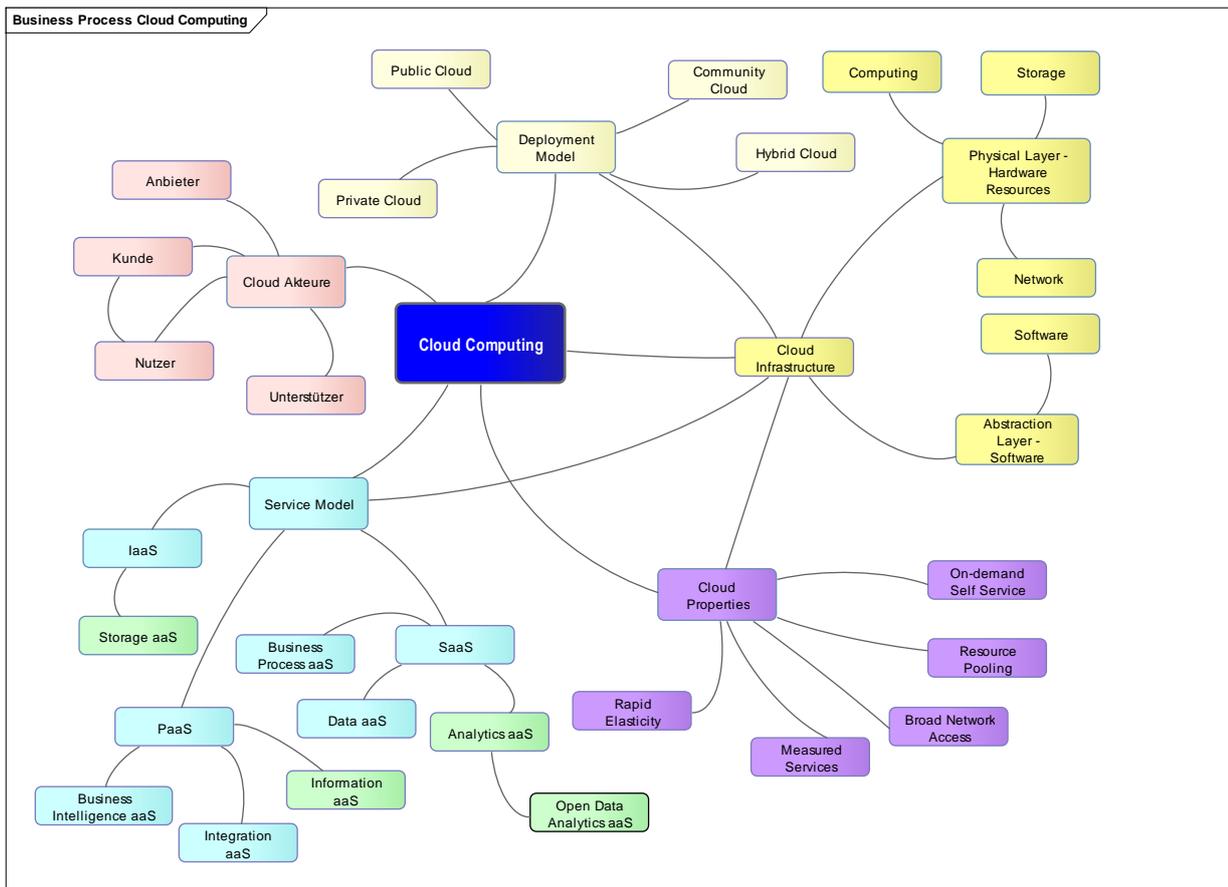


Abbildung 1: Beziehungen zwischen den Begrifflichkeiten im Cloud Computing

3.4 NIST Cloud Service Taxonomie

Eine der Zusammenstellung von Cloud-Diensten, die auf die einzelnen Cloud-Dienstmodelle aufgeteilt wurden, ist in der in

Abbildung 2 dargestellten Cloud-Dienste Taxonomie von NIST zu finden, die als Bestandteil des NIST Referenzmodells (NIST - National Institute of Standards and Technology, 2011) veröffentlicht wurde. Interessant ist, dass die Bereitstellung von Speicher, wie sie als Teil eines Integrations- und Analyseprozesses erforderlich ist, bereits als Spezialfall von IaaS vorgesehen ist. Business Intelligence (BI)-Plattformen und Integrationsplattformen werden als Spezialfälle von PaaS betrachtet. AaaS ist in der Taxonomie noch nicht vorgesehen. Diese Spezialisierungen sind jedoch in Abbildung 1 bereits berücksichtigt.

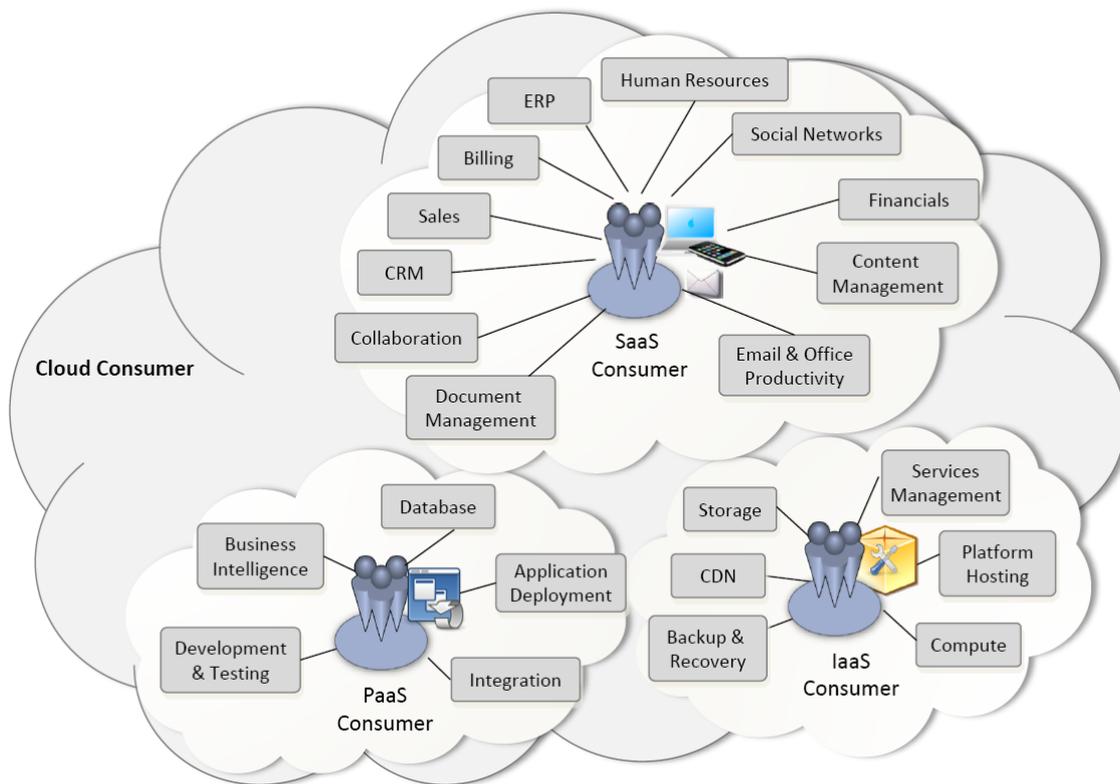


Abbildung 2: NIST Cloud-Dienste Taxonomie²³

3.5 NIST Cloud-Definition

Die vom NIST vorgeschlagene Definition von Cloud Computing (Mell & Grance, The NIST Definition of Cloud Computing (Draft), 2011) lehnt sich stark an die oben eingeführten Begriffe an. Ihr Fokus liegt auf der einfachen und möglichst automatischen Bereitstellung von Ressourcen:

„Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.“

3.6 BSI Cloud-Definition

Das BSI hat in (BSI - Bundesamt für Sicherheit in der Informationstechnik, 2011) eine deutsche Version dieser Definition vorgestellt, die sich stärker an wirtschaftlichen Aspekten und der Existenz wohldefinierter Schnittstellen orientiert. Die Definition adressiert somit implizit Interoperabilitätsaspekte, ohne jedoch auf konkrete Spezifikationen von Cloud-Schnittstellen und Protokollen zu verweisen.

²³ Vgl. (NIST - National Institute of Standards and Technology, 2011), Abb. 6, Seite 6

Cloud Computing bezeichnet das dynamisch an den Bedarf angepasste Anbieten, Nutzen und Abrechnen von IT-Dienstleistungen über ein Netz. Angebot und Nutzung dieser Dienstleistungen erfolgen dabei ausschließlich über definierte technische Schnittstellen und Protokolle. Die Spannweite der im Rahmen von Cloud Computing angebotenen Dienstleistungen umfasst das komplette Spektrum der Informationstechnik und beinhaltet unter anderem Infrastruktur (z.B. Rechenleistung, Speicherplatz), Plattformen und Software.

3.7 NIST Akteure

NIST identifiziert in seiner Cloud-Referenzarchitektur (NIST - National Institute of Standards and Technology, 2010) eine große Anzahl von Akteuren, die sich jedoch alle in das grobe Schema Anbieter, Kunden und Unterstützer von Cloud-Diensten eingliedern. NIST Akteure können durch Personen, Organisationen und technische Systeme repräsentiert werden.

- **Anonymer Nutzer:** Ein nicht identifizierter und authentisierter Akteur, der Cloud-Dienste nutzt.
- **Kunde:** Eine Person oder Organisation, die in einer Geschäftsbeziehung mit einem Cloud-Anbieter steht.
- **Nutzer:** Kunden und von ihnen berechnigte Endnutzer eines Cloud-Dienstes.
- **Kundenadministrator:** Spezieller Nutzer eines Kunden, der administrative Aufgaben für den Kunden ausübt.
- **Cloud Nutzer:** Eine Person, die Cloud-Dienste benutzt aber keine geschäftliche Beziehung zu deren Anbieter besitzt.
- **Payment Broker:** Eine unterstützender Akteure, der finanzielle Transaktionen zwischen Kunden und Anbietern von Cloud-Diensten abwickelt.
- **Dienstanbieter:** Eine Organisation, die Cloud-Dienste anbietet, überwacht und abrechnet.
- **Transportagent:** Eine Organisation, die die für den Betrieb einer Cloud erforderlichen Ressourcen transportiert. („A business organization that provides physical transport of storage media such as high-capacity hard drives“).
- **Rechtsbehörde:** Gerichte, Behörden oder Polizei.
- **Identity Provider:** Eine Einrichtung, die digitale Identitäten für Personen, Organisationen oder Hard/Software bereitstellt und pflegt.
- **Cloud Dienst Broker:** Dienstanbieter, der Cloud-Dienste von unterschiedlichen Dienstanbietern zwischen diesen und an weitere Kunden vermittelt.

In Bezug auf die betrachtete Fragestellung sind folgende Akteure von besonderem Interesse:

- **Anonyme Nutzer** von AaaS erlauben die Nutzung von Analysediensten, ohne Informationen über ihre Identität preisgeben zu müssen. Dies ist insbesondere im Umfeld offener Daten ein wichtiges Kriterium, da die Daten selber auch anonym genutzt werden können.
- **Dienstanbieter** sind öffentliche oder privatwirtschaftliche Einrichtungen, die Analysedienste für eigene Zwecke nutzen oder diese unentgeltlich bzw. kommerziell anbieten. In der vorliegenden Studie wird unterschieden, ob der AaaS-Dienstanbieter gleichzeitig auch Anbieter der zu analysierenden Daten ist, ggf. im Sinne von Storage as a Service oder Data as a Service, oder ob diese Rolle außerhalb der betrachteten Cloud-Akteure liegt.
- **Dienstnutzer/-kunden** sind natürliche oder juristische Personen, die AaaS nutzen. Während Bürger im Allgemeinen beide Rollen vereinen, findet bei Unternehmen eine Trennung in Kunden und Nutzer statt. Das Unternehmen selbst agiert als Kunde, der Mitarbeiter als Nutzer des AaaS.

3.8 NIST Anwendungsfälle

NIST hat seine Anwendungsfälle (englisch: use cases) (NIST - National Institute of Standards and Technology, 2010) mit dem Ziel entwickelt, folgende Fragestellungen zu beantworten:

- **Interoperabilität:** Kann ein Cloud-Kunde Dienste von verschiedenen Cloud-Anbietern in einem gemeinsamen Anwendungskontext nutzen? Kann der Kunde die administrativen und geschäftlichen Interaktionen mit verschiedenen Cloud-Anbietern einheitlich durchführen?
- **Portabilität:** Kann ein Cloud-Kunde den Cloud-Anbieter wechseln, ohne dabei technische und administrative Änderungen durchführen zu müssen?
- **Sicherheit:** Welche Unterstützung bietet ein Cloud-Anbieter, um in der Cloud gespeicherte Daten jederzeit verfügbar zu halten und unberechtigte Zugriffe auszuschließen?

Da eine vollständige, standardbasierte Antwort auf diese Fragestellungen kurzfristig nicht gegeben werden konnte, erarbeitete NIST im „Standards Acceleration to Jumpstart Adoption of Cloud Computing (SAJACC)“ Projekt (NIST - National Institute of Standards and Technology, 2010) pragmatische Lösungen um die Portabilität, Interoperabilität und Sicherheit im Cloud Computing zu erhöhen. Die NIST-Anwendungsfälle sind in (NIST - National Institute of Standards and Technology, 2010) beschrieben und online unter (NIST - National Institute of Standards and Technology, 2010) verfügbar. Sie unterteilen sich in die drei Themenbereiche:

- **Cloud Management**
- **Cloud-Interoperabilität**
- **Cloud-Sicherheit**

Cloud Management

Der Themenbereich Cloud Management umfasst administrative Anwendungsfälle und die Portabilität von Daten:

- **Open an account:** Dienstleister eröffnet ein Konto für einen neuen Kunden.
- **Close an account:** Dienstleister schließt das Konto eines Kunden.
- **Terminate an account:** Dienstleister löscht das Konto eines Kunden.
- **Copy data objects into a cloud:** Kunde kopiert seine lokalen Daten in die Cloud-Infrastruktur.
- **Copy data objects out of a cloud:** Kunde kopiert seine Daten aus der Cloud-Infrastruktur in seine lokale Umgebung.
- **Erase data objects in a cloud:** Daten in der Cloud-Infrastruktur werden auf Initiative des Kunden gelöscht.
- **VM control, allocate VM instance:** Kunde erzeugt virtuelle Maschinen (VM, englisch: Virtual Machine), die seinen Anforderungen bezüglich Durchsatz und Sicherheit genügen.
- **VM control, manage virtual machine instance state:** Kunde startet, stoppt, beendet oder ändert den Status seiner VM
- **Query cloud provider capabilities and capacities:** Angebotsverhandlungen zwischen Cloud Anbietern und Cloud Kunden.

Im Kontext von AaaS sind die ersten sechs Anwendungsfälle von Interesse. Hier ist zu untersuchen, inwieweit ODAaaS-Werkzeuge diese Anwendungsfälle implementieren müssen.

Cloud-Interoperabilität

Der Themenbereich Cloud-Interoperabilität umfasst das Zusammenspiel von Diensten verschiedener Cloud-Anbieter und ist für Fragestellungen bezüglich Portabilität und Interoperabilität wichtig:

- **Copy data objects between cloud providers:** Datenportabilität – der Kunde veranlasst den Transfer von Daten zwischen verschiedenen Diensteanbietern.
- **Dynamic operation dispatch to IaaS clouds:** IaaS-Dienst-Interoperabilität – der Kunde ermittelt regelgesteuert die für seine Zwecke geeignetste Cloud-Infrastruktur (IaaS) und nutzt diese entsprechend seinen Anforderungen.
- **Cloud burst from data center to cloud:** IaaS-Interoperabilität – dynamische Nutzung und Freigabe von Cloud IaaS-Ressourcen, die von ggf. mehreren Cloud-Anbietern zur Verfügung gestellt werden.
- **Migrate a queuing-based application:** Dienstportabilität – Migration eines existierenden Publish-/Subscribe- oder sonstigen nachrichtenbasierten Dienstes zwischen Cloud-Anbietern.
- **Migrate (fully stopped) VMs from one cloud provider to another:** VM-Portabilität – Nahtlose Migration gestoppter VMs zwischen verschiedenen IaaS-Diensteanbietern.

Im Kontext von AaaS sind diese Anwendungsfälle von geringerem Interesse, da sie sich weitestgehend mit IaaS-Interoperabilität beschäftigen. Das Kopieren offener Daten zwischen verschiedenen Cloud-Infrastrukturen und Cloud-burst Szenarien können zwar auch bei AaaS auftreten, verursachen jedoch keine spezifischen Anforderungen.

Cloud-Sicherheit

Der Themenbereich Cloud-Sicherheit umfasst Fragen der Sicherheit, insbesondere der Authentisierung und Autorisierung.

- **User account provisioning:** Der Kunde beantragt die Einrichtung eines Kontos. Dabei soll möglichst eine Synchronisation mit bereits existierenden Identitäten und Berechtigungen erfolgen.
- **User authentication in the cloud:** Der Kunde authentisiert sich gegenüber der Cloud, möglichst unter Nutzung standardisierter Verfahren wie SAML, OpenID oder Kerberos.
- **Data access authorization policy management in the cloud:** Der Kundenadministrator verwaltet die Berechtigungen (englisch: policies) zum Zugriff auf in der Cloud gespeicherte Daten (schreiben, lesen, ändern).
- **User credential synchronization between enterprises and the cloud:** Der Kunde veranlasst Anpassungen zwischen den Identifikations- und Berechtigungssystemen beim Kunden und dem Cloud Anbieter.
- **eDiscovery:** Pflege und Verwaltung von Daten und Metadaten (Logs, Zugriffsprotokollen) in der Cloud, um rechtlichen Anforderungen zu genügen.
- **Security Monitoring:** Automatische Überwachungen der Infrastruktur des Diensteanbieters zum Nachweis der Einhaltung getroffener Vereinbarungen und rechtlicher Regelungen.
- **Sharing of access to data in a cloud:** Erteilung der Berechtigung für Kunden auf Daten anderer Kunden zuzugreifen.

Im Kontext von AaaS ist der siebente Anwendungsfall von Interesse. Sofern im Rahmen der Datenveredlung durch AaaS neue Erkenntnisse, repräsentiert durch aggregierte und vorausgewertete Daten, gewonnen werden können stellt sich die Frage, inwieweit diese „Zwischenergebnisse“ auch anderen Nutzern zur Verfügung gestellt werden können oder müssen.

3.9 NIST Referenzarchitektur

Anerkannte, herstellerunabhängige, stabile Cloud-Architekturen sind außer dem NIST-Ansatz nicht verbreitet. Unterschiedliche Organisationen arbeiten an Cloud-Referenzarchitekturen, die jedoch nach ihrer Verabschiedung noch einer Konsolidierung durch die Standardisierung bedürfen. Im Folgenden wird daher ausschließlich der NIST-Ansatz vorgestellt. Ziel ist es einerseits, die große Nähe von Cloud-Architekturen und SOA aufzuzeigen und die Cloud-spezifischen Architektur Aspekte deutlich hervorzuheben. Andererseits sollen diejenigen Schnittstellen zwischen der Cloud und ihren Nutzern bzw. zwischen Cloud-Diensten und Cloud-Infrastrukturen identifiziert werden, die relevant für Interoperabilitätsaspekte sind. Einen Überblick über Open Source Cloud-Architekturen gibt NIST in (NIST - National Institute of Standards and Technology, 2011).

Die in Abbildung 3 dargestellte NIST Reference Architecture (NIST - National Institute of Standards and Technology, 2011) (vgl. Abb. 1, Seite 3) basiert auf den oben vorgestellten NIST-Rollen. Aus dieser Abhängigkeit lassen sich sehr gut die Beziehungen zwischen Architektur, Anwendungsfällen und Rollen/Akteuren erkennen.

Für ODAaaS-Werkzeuge ist zu ermitteln, ob und wie sie sich in ein derartiges Referenzmodell einordnen lassen. Soweit ODAaaS als Spezialfall von SaaS betrachtet werden kann, könnte ein Cloud-Anbieter zugehörige Werkzeuge in seine Cloud-Infrastruktur einbetten, die ihm die gesamte Funktionalität der Cloud ohne weiteren Implementierungsaufwand bereitstellt. Werden die Werkzeuge dagegen „nur“ extern bereitgestellt und über Web-Schnittstellen zugreifbar gemacht, muss die benötigte Funktionalität einer Cloud-Infrastruktur explizit nachgebildet werden.

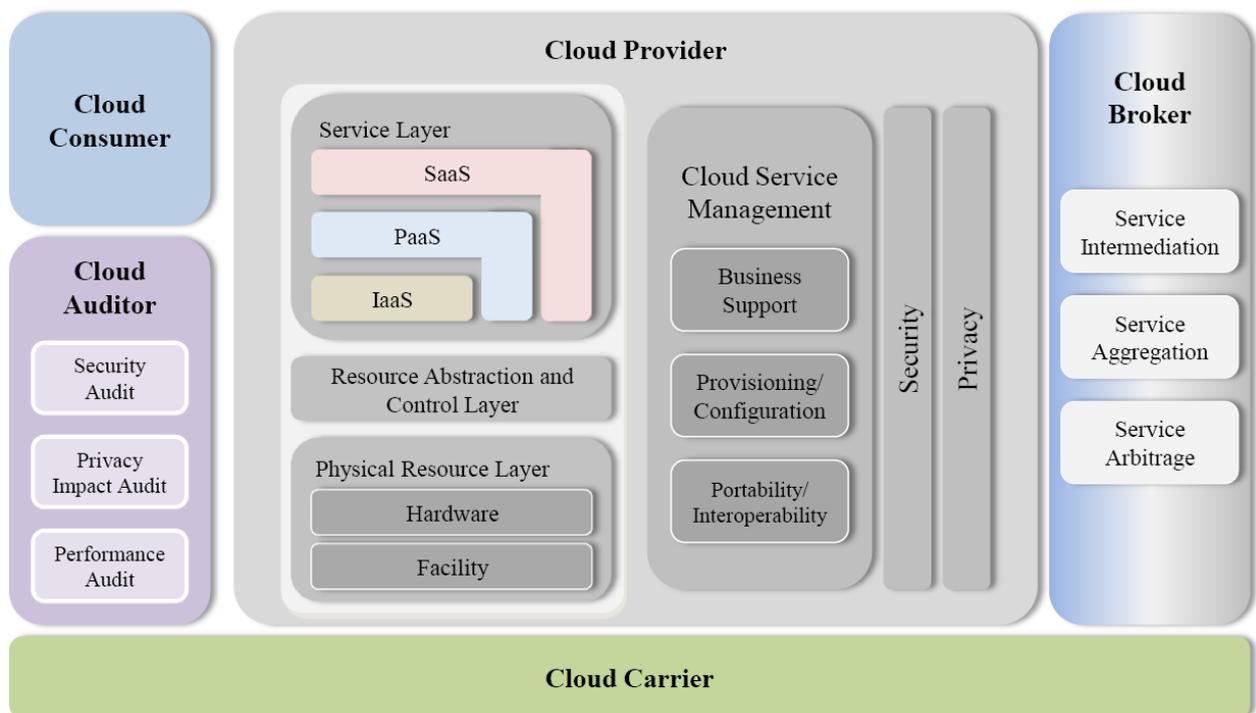


Abbildung 3: NIST Cloud-Referenzarchitektur²³

Die Rollen oder Akteure bezeichnen die mit der Bereitstellung und Nutzung von Cloud-Diensten beschäftigten Personen und Organisationen. Jede Rolle führt spezifische Interaktionen mit der Cloud aus, die durch zugehörige Anwendungsfälle beschrieben sind. Die Cloud-Architektur definiert Komponenten und Schnittstellen, über die die Akteure die Anwendungsfälle ausführen können.

Interoperabilität und Portabilität

Die vorgestellte Architektur identifiziert eine Reihe von für Einbettung, Interoperabilität und Portabilität wichtigen Schnittstellen. Dabei werden verschiedene Typen von Schnittstellen unterschieden:

- Management Schnittstellen zu Entwicklern, Betreibern und Kunden
- SaaS Schnittstellen zu PaaS und ggf. IaaS-Funktionen
- PaaS Schnittstellen zu IaaS Funktionen

Der erste Typ von Schnittstellen ist relevant für Interoperabilitätsfragen. Dabei wird unter der Interoperabilität von zwei oder mehr Cloud Computing-Plattformen deren Fähigkeit verstanden, über standardisierte Schnittstellen zusammenarbeiten zu können. Plattform A tritt dabei beispielsweise als Kunde von Plattform B auf, um Cloud Burst-Szenarien zu realisieren, Daten zu migrieren, virtuelle Maschinen zu migrieren, Nutzungs- und Zugriffsrechte zu übertragen und ähnliches. Interoperabilitätsfragen auf Anwendungsebene, d.h. zwischen SaaS Dienstangeboten werden dagegen nicht betrachtet, da sie nicht Cloud spezifisch sind.

Der zweite und dritte Typ von Schnittstellen ist relevant für Portabilitätsfragen. Dabei wird unter der Portabilität von Cloud-Diensten deren Fähigkeit verstanden, auf unterschiedlichen Cloud Computing-Plattformen ausgeführt werden zu können. Diese Eigenschaft ist erforderlich, wenn Anwendungen oder Daten zwischen unterschiedlichen Cloud-Anbietern migriert werden sollen. Ohne Portabilität besteht für den Kunden die Gefahr der festen Bindung an einen Cloud-Anbieter, das sogenannte „Vendor Lock-In“.

Die an dieser Stelle verwendete Interpretation der Begriffe Portabilität und Interoperabilität deckt sich mit deren Verwendung im Eckpunktepapier des BSI (BSI - Bundesamt für Sicherheit in der Informationstechnik, 2011).

Auf Infrastrukturebene existieren derzeit bereits standardisierte Schnittstellen, wie das *Open Cloud Computing Interface* (OGF - Open Grid Forum, 2011), (OGF - Open Grid Forum, 2011) des Global Grid Forums oder das Open Virtualization Format (OVF) der DMTF (DMTF - Distributed Management Task Force, 2010).

Das NIST verwendet in der ersten Version seiner Cloud-Referenzarchitektur (NIST - National Institute of Standards and Technology, 2011) die Begriffe Portabilität und Interoperabilität in Anlehnung an existierende Standards wie folgt:

- **Portabilität** bezeichnet
 - die Fähigkeit, Daten von einem System auf ein anderes System zu transferieren, ohne Beschreibungen der Daten und verarbeitende Systeme anpassen zu müssen,
 - die Eigenschaft von Software, auf unterschiedlichen Rechnern und Betriebssystemen ablauffähig zu sein.
- **Interoperabilität** bezeichnet die Fähigkeit von Systemen unter festgelegten Bedingungen kommunizieren und Daten austauschen zu können.

Aus diesen Interpretationen leitet NIST folgende Aufgaben für die Anbieter von Cloud-Diensten ab:

- **Datenportabilität:** Daten müssen in die Cloud hinein und aus der Cloud heraus kopiert werden können. Der Transfer großer Mengen von Daten muss möglich sein.
- **Dienstinteroperabilität:** Daten und Anwendungen müssen für Cloud-Nutzer in unterschiedlichen Cloud-Umgebungen über die gleichen Management-Schnittstellen nutzbar sein.
- **Systeminteroperabilität:** Gestoppte virtuelle Maschinen sowie Anwendungen und Diensten müssen mit ihren Daten zwischen verschiedenen Cloud Umgebungen migriert werden können.

Dieser Ansatz weicht in der Nutzung der Begriffe von der vorherigen ab, beschreibt aber dennoch die gleichen Effekte. Eine verbindliche Interpretation der beiden Begriffe im Kontext Cloud Computing existiert derzeit nicht.

Im Kontext von AaaS sind insbesondere die Anforderungen Daten-Portabilität und aus Sicht des AaaS-Anbieters Software-Portabilität bzw. Dienst-Interoperabilität wichtig. Offene und aus diesen gewonnene, veredelte Daten müssen aus der AaaS-Cloud-Infrastruktur heraus zugreifbar sein und aus dieser herauskopiert werden können. Angebotene AaaS-Werkzeuge müssen in unterschiedlichen Cloud-Infrastrukturen ablauffähig und auf gleiche Art und Weise nutzbar sein.

4 Bestandteile von ODAaaS

Im folgenden Kapitel sollen Spezifika von ODAaaS analysiert und relevante Anforderungsdimensionen abgeleitet werden. Neben technischen Anforderungen werden auch organisatorische und wirtschaftliche Anforderungen betrachtet, um eine möglichst ganzheitliche Bewertung und Abschätzung zukünftiger Entwicklungslinien für umfassende Lösungen zu erlauben.

Unter ODAaaS wird im Rahmen dieser Studie das Angebot eines Verfahrens verstanden, das offene und nicht offene Daten unter Verwendung einer cloudbasierten Infrastruktur aufbereitet, anreichert und durch statistische Verfahren erschließt. Die offenen Daten müssen dabei im Rahmen einer Open Data-Plattform referenziert, über Metadaten erschlossen und durch einen Datenbereitsteller zur Verfügung gestellt werden. Die Verschränkung offener und nicht offener Daten und die hiermit verbundenen Potentiale werden dabei bewusst mit eingeschlossen.

Im Folgenden soll auf Basis eines einfachen Prozessmodells eine Grundlage für ODAaaS systematisiert werden. Hierbei werden vier sequentielle **Prozessschritte** angenommen, welche durchlaufen werden müssen, um ODAaaS als Gesamtprozess abbildbar zu machen:

- **Retrieve** - Abruf von Daten
- **Revise** - Ausgestaltung und Bereinigung von Daten
- **Integrate** - Daten miteinander in Beziehung setzen
- **Analyze** - Daten analysieren

Für alle Schritte bestehen, wie eingangs erwähnt, unterschiedliche Anforderungen. Es ist zu beachten, dass die einzelnen Anforderungen nur in wenigen Fällen idealtypisch einem einzelnen Schritt zugeordnet werden können, sondern in einigen Fällen Querbezüge zwischen den Schritten relevant sind.

Bei den betrachteten **Anforderungsdimensionen** handelt es sich um:

- **Technische** Anforderungen
- **Organisatorische** Anforderungen
- **Wirtschaftliche** Anforderungen

Als Basis für die **technischen Anforderungen** an ODAaaS werden die in Kapitel 2 und die in Kapitel 3 genannten Grundprinzipien von Open Data und Cloud Computing betrachtet. Zusätzlich ergeben sich aus dem Kontext des Anwendungsgebiets weitere Aspekte, bezüglich der Integrierbarkeit und Zusammenführbarkeit von Daten zum Zweck statistischer Auswertungen. Dies gilt insbesondere, wenn sich Analysen nicht auf einzelne Variablen, sondern auf multivariate Analyseverfahren²⁴ beziehen.

Organisatorische Anforderungen ergeben sich aus organisatorischen Prozessen an einem der Endpunkte einer Analyse, z.B. bei der Bereitstellung von Daten, Datenkatalogen und Analysefunktionalitäten. Es können Querbezüge zu technischen Anforderungen bestehen, insofern organisatorische Anforderungen möglicherweise durch den Einsatz geeigneter Technologien reduziert werden können. Hauptsächlich sollen Anforderungen berücksichtigt werden, die einen direkten Prozessbezug haben, d.h. sich auf die Interaktion von Akteuren beziehen oder auf organisatorische Bedingungen zur Bereitstellung der benötigten Artefakte.

Wirtschaftliche Anforderungen beziehen sich auf die Vermarktung von Analysewerkzeugen oder die Erbringung von Dienstleistungen auf Basis der Verwendung von Analysewerkzeugen im Kontext offener Daten.

²⁴ Gleichzeitige Analyse mehrere Variablen

Hierbei können auch Querbeziehungen zu technischen Anforderungen eine Rolle spielen, sofern sich aus technischen Funktionalitäten verbesserte Marktchancen ergeben.

Eine Abgrenzung allgemeiner Datenanalyse- zu Open Data Analytics-Konzepten muss sich dabei an den Spezifika der Rahmenbedingung von Open Data-Plattformen orientieren und deren Leistungsumfang mit einbeziehen. Übergeordnetes Ziel ist die Ableitung von Anforderungen, die den spezifischen Charakter von Open Data-Konzepten und die daraus entstehenden Erfordernisse an eine Gesamtkonzeption von ODAaaS untersuchbar machen. Als typische Open Data-Plattform wird die in Kapitel 2.4 beschriebene CKAN-Basissoftware als Referenz herangezogen. Obwohl CKAN nicht flächendeckend im Einsatz ist und nicht alle Funktionen genutzt werden, ist auf Basis von CKAN eine sinnvolle Ableitung der Anforderungen an ODAaaS-Konzepte möglich. Entsprechende empirisch beobachtbare Einschränkungen, die sich aus der Verwendungspraxis einzelner Funktionalitäten der CKAN-Plattform ergeben, werden in Kapitel 7 unter Bezug auf mögliche Ausgestaltungsformen von ODAaaS adressiert.

Ein weiterer wesentlicher Aspekt von ODAaaS besteht in der Bereitstellung von Funktionalitäten im Rahmen eines cloudbasierten Diensts. Dabei kommen die in Abbildung 1 dargestellten Servicetypen in Frage:

- **Storage as a Service:** Angebot von Speicherinfrastruktur - Spezialfall von IaaS, der es ermöglicht, offene oder bereits veredelte Nutzerdaten zu speichern und insbesondere Ergebnisse nachgelagerter Prozessschritte aufzubewahren.
- **Information as a Service (Typ 1):** Zugriff auf aufbereitete Daten - Spezialfall von SaaS, der es ermöglicht, Nutzern offene Daten in geeigneter Form zugänglich zu machen und eine Vorverarbeitung erlaubt. Dazu gehört, nach Nutzervorstellungen zu filtern, unterschiedliche Formate zur Verfügung zu stellen, auf Qualität zu prüfen, unterschiedliche Datenbestände zusammenzuführen oder auf höhere Abstraktionsebenen umzurechnen.
- **Information as a Service (Typ 2):** Angebot einer Laufzeitplattform für die Analyse aufbereiteter Daten - Spezialfall von PaaS, der Entwicklungs- und Laufzeitumgebungen bereitstellt, in denen Nutzer eigene Algorithmen ausführen können, um z.B. Datenveredelung durchzuführen.
- **Analytics as a Service:** Angebot von Analyse-Software - Spezialfall von SaaS, der es ermöglicht, dem Nutzer eine Benutzeroberfläche zur Verfügung zu stellen, die es erlaubt, Analysen durchzuführen und deren Ergebnisse in unterschiedlicher Form darzustellen.

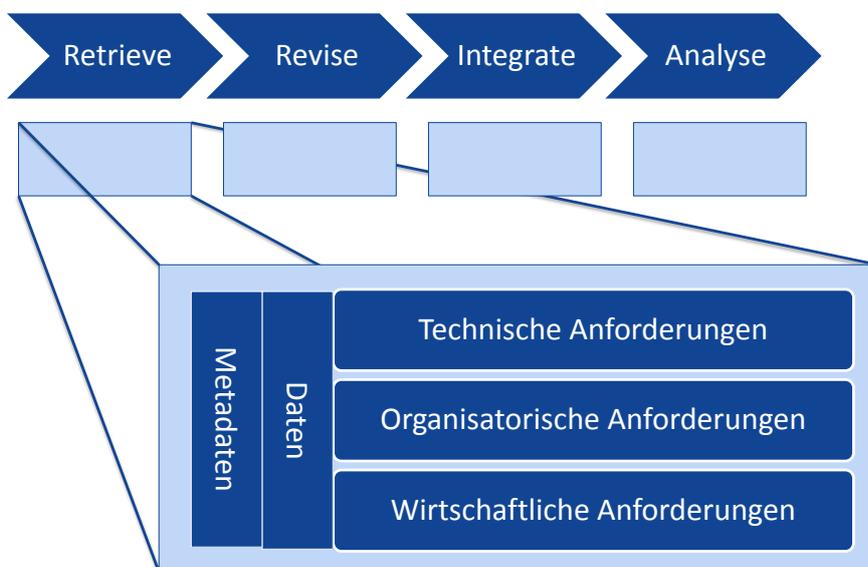


Abbildung 4: Prozessschritte und zugehörige Anforderungen in ODAaaS

Aus den Anforderungen ergibt sich für ODAAaaS das in Abbildung 4 dargestellte Analyseraster. Es dient sowohl der Betrachtung einzelner Schritte und spezifischer Funktionalitäten (Retrieve, Revise, Integrate, Analyze) als auch der Betrachtung der in diesen Schritten ableitbaren Anforderungen (technisch, organisatorisch, wirtschaftlich). Dabei werden die Anforderungen an Daten und an die beschreibenden Metadaten betrachtet.

4.1 Datenermittlung - Retrieve

Im Retrieve-Schritt werden offene Daten, die durch Datenbereitsteller hinterlegt, durch Metadaten beschrieben und – z.B. durch eine Open Data-Plattform – referenziert sind, durch ein IT-System abgerufen, eingelesen und somit für nachfolgende Verarbeitungsschritte verfügbar gemacht. Sie werden im Retrieve-Vorgang weder inhaltlich erschlossen noch integriert. Daten müssen dazu über durch Metadaten beschriebene Zugriffspunkte bezogen und eingelesen werden.

Ein zu berücksichtigender Aspekt für ODAAaaS besteht im Retrieve-Schritt darin, dass offene Daten nur selten über einen direkten Datenbankzugriff bzw. über Schnittstellen abrufbar sind. In der heutigen Praxis werden sie oft als Dateien über eine URL von unterschiedlichen Datenbereitstellern dezentral zur Verfügung gestellt. ODAAaaS-Konzepte müssen daher auf Datenbeständen aufbauen, die nicht per se in Datenbanken oder Data Warehouses abgelegt sind. Bei der Bereitstellung und Aktualisierung offener Daten ist nicht gewährleistet, dass diese in einheitlicher Weise und in einheitlichem Format publiziert werden. Im Extremfall müssen mit jeder Datenaktualisierung Routinen zur Datenbehandlung ausgeführt werden.

Unterstützung von Schnittstellen und Identifikation von Zugriffspunkten aus Metadaten

Wie bereits dargestellt, werden offene Daten nicht notwendigerweise direkt zentral auf einer Open Data-Plattform vorgehalten. Oft existiert nur ein Katalog von Metadaten zu den verteilt vorliegenden Datensätzen. ODAAaaS-Lösungen müssen daher Komponenten beinhalten, die standardisiert das Auslesen von Metadaten aus einer Open Data-Plattform erlauben und den Zugriff auf die verteilt hinterlegten offenen Daten gewährleisten. Für ODAAaaS ist diese Anforderung unkritisch, da Daten als Dateien oder mitunter auch über webbasierte APIs über das Internet bezogen werden können. Spezielle Datenbank-Konnektoren sind nur in Einzelfällen erforderlich. Importroutinen für hinterlegte Metadaten und Dateiformate müssen jedoch verfügbar sein. Sie weisen unterschiedliche Komplexität auf, z.B. für CSV-Daten oder eingebettete Datentabellen im PDF-Format.

Für den Retrieve-Schritt von ODAAaaS ist daher nur die Verarbeitung von Metadaten notwendig, die sich auf den Zugriffspunkt eines Datensatzes beziehen. Diese Metadaten können bei vielen Open Data-Plattformen vollständig über eine webbasierte Schnittstelle abgefragt werden.

Eine organisatorische Anforderung auf Seiten der Datenbereitstellung ergibt sich aus der Qualität der Metadaten und die Aktualität der Daten. Diese müssen durch den Datenbereitsteller gewährleistet werden. Anderenfalls müssen nachgelagerte ODAAaaS-Prozessschritte diese Unzulänglichkeit behandeln.

Bei der Einbeziehung privater Daten in den Analyseprozess gelten vergleichbare Aussagen. Bei privaten Daten kann nicht von der Existenz standardisierter Metadaten ausgegangen werden.

Aktualität und Verfügbarkeit von Daten

Die Aktualität von Daten und deren Verfügbarkeit sind sowohl aus organisatorischer als auch aus wirtschaftlicher und technischer Sicht relevant. In Bezug auf die Realisierung wirtschaftlicher Potentiale spielt die Aktualität von Daten eine tragende Rolle. Hiervon sind sowohl die Ebene der Datenbereitstellung, die Open Data-Plattform und mittelbar auch die Möglichkeiten werthaltiger Analysen im Rahmen von ODAAaaS betroffen. Daten dürfen nicht

grundsätzlich als statisch angesehen werden. Im Rahmen von ODAaaS müssen daher entsprechende Funktionen vorgehalten werden, die in regelmäßigen Abständen die Metadaten einer Open Data-Plattform auslesen, auf aktualisierte Datensätze prüfen und die Verfügbarkeit auf Seiten der Datenbereitsteller feststellen. Alternativ kann dies bereits durch organisatorische und technische Maßnahmen einer Open Data-Plattform gewährleistet werden.

Die Verwendung von Link Checker-Funktionalitäten ist ein leicht umsetzbarer technischer Lösungsansatz. Dennoch müssen an entsprechende Funktionalitäten organisatorische und technische Prozesse gekoppelt werden, die sich auch daran orientieren, in welcher Weise Daten im Rahmen von ODAaaS hinterlegt werden.

Aus organisatorischer Sicht ist zu beachten, dass Datenbereitsteller nicht zwingend selbst Daten in ein Open Data-Portal einpflegen. Dies kann auch im Rahmen des Aufbaus oder Betriebs von Open Data-Portalen durch Dritte geschehen. Unklare Prozessketten und Kommunikationswege können Einfluss auf die Aktualität und Verfügbarkeit von Daten haben. Dies ist z.B. der Fall, wenn Datenbereitsteller keine Kenntnis über die Referenzierung „ihrer“ offenen Daten in einer Open Data-Plattform haben.

Grundsätzlich ist zu berücksichtigen, dass verschiedene Open Data-Lizenzen, wie z.B. die „Datenlizenz Deutschland“²⁵, die dauerhafte Verfügbarkeit der Daten nicht garantieren.

Umfang der Datenkataloge

Der Umfang verfügbarer Daten wird für die wirtschaftlichen Erfolgsaussichten von ODAaaS eine zentrale Rolle spielen und stellt eine organisatorische Anforderung dar. Der Umfang verfügbarer Datenkataloge und Daten hat einen wesentlichen Effekt auf die Attraktivität von ODAaaS für den Endnutzer. Die Attraktivität von ODAaaS-Angeboten wird daher stark durch die Datenbereitsteller geprägt.

Umfang und Beschaffenheit der Daten beeinflussen die Schritte Retrieve und Revise, was wiederum Auswirkungen auf die wirtschaftliche Verwertbarkeit eines ODAaaS-Dienstes hat. Der Umfang der Datenkataloge kann sich dabei auch auf Derivate offener Daten beziehen. Bei entsprechendem Nutzerinteresse können z.B. offene Daten vorgefiltert und Teildatensätze separat ausgewiesen werden. Wird ein ODAaaS-Dienst entsprechend des Nutzerinteresses aufgebaut, entstehen daraus wiederum technische und organisatorische Anforderungen für die Bereitstellung der Daten. Das Nutzerinteresse kann über eine Analyse des Nutzerverhaltens ermittelt und über die technische Filterung von Datensätzen berücksichtigt werden.

Zugriff auf offene und veredelte Daten mittels Cloud-Diensten

Für den effizienten Zugriff auf Daten und ihre weitere Verwertung sollten die Daten in geeigneter Weise gespeichert werden, z.B. durch die zentrale Nutzung von „Storage as a Service“. Auch für die Analyse von Daten ist eine solche zentrale, Cloud-basierte Vorhaltung hinsichtlich Verfügbarkeit und Performanz des Datenzugriffs vorteilhaft. Die Bereitstellung eines zentralen Datenspeichers als Teil einer Open Data-Plattform, der durch Cloud Storage-Dienste realisiert werden kann, ist insbesondere dann sinnvoll, wenn mittels entsprechender Werkzeuge aufbereitete Daten wiederum Dritten – im Sinne von Datenwertschöpfungsketten – für die weitere Nutzung und Bearbeitung zugänglich gemacht werden sollen.

Da es sich um ohnehin offene Daten handelt, sind Datenschutzfragen vor dem Hintergrund der aktuellen Diskussion der Vertrauenswürdigkeit der Anbieter von cloudbasierten Datenspeicherdiensten als unkritisch zu bewerten. Wenn ein solcher zentraler Datenspeicher nicht nur von Datennutzern für die Speicherung von – im Rahmen der Aufbereitung oder Analyse von Rohdaten erzeugten – Datenderivaten, sondern auch von den Datenbereitstellern (i.d.R. Verwaltungen) für die Publikation der originären offenen Daten verwendet wird, können jedoch Konsistenzprobleme durch redundante Datenhaltung entstehen (z.B. Publikation auf

²⁵ Vgl. <https://www.govdata.de/lizenzen>

Verwaltungswebseiten und im Datenspeicher einer Open Data-Plattform). Derartigen Konsistenzproblemen kann durch geeignete organisatorische und technische Maßnahmen begegnet werden.

Im Allgemeinen wird eine Speicherung von replizierten Rohdaten im Retrieve-Schritt jedoch nicht erfolgen. Eine Speicherung ist erst als Ergebnis der Aufbereitung (Revise) oder Zusammenführung (Integrate) erforderlich.

Zusammenfassung

Zusammengefasst lassen sich für den Retrieve-Schritt folgende Aussagen treffen. Im Retrieve-Schritt müssen

- Daten über Schnittstellen zugänglich und Zugriffspunkte bekannt sein;
- Daten möglichst aktuell bereitgestellt werden;
- Daten im Einzelfall in einer zentralen Infrastruktur repliziert hinterlegt werden. Dies hat Auswirkungen auf die Freiheitsgrade bei der Auswahl eines Ansatzes für die Integration von Daten.

4.2 Datenbereinigung - Revise

Im Revise-Schritt werden Daten so prozessiert, dass sie integriert und für einfache Anwendungen von Analysewerkzeugen nutzbar gemacht werden können. Am Ende dieses Schritts sind offene Daten inhaltlich über dazugehörige Metadaten erschlossen und stehen, ggf. konvertiert, für spätere Schritte in ausreichender Qualität zur Verfügung. Dies impliziert, dass entsprechende Datenkonvertierungs-, -bereinigungs- und Qualitätsprüfungsprozesse implementiert sind. Während dieses Vorgangs werden Datensätze jedoch noch nicht integriert, so dass Analysen sich weiterhin nur auf einzelne Datensätze aus einer gemeinsamen Datenquelle beziehen können.

Inhaltliche Einordnung von Daten auf Basis von Metadaten

Metadaten werden von Open Data-Plattformen bereitgestellt und sind für nachfolgende Verarbeitungsschritte relevant, da nur durch sie eine korrekte Interpretation und Rückverfolgung zur Datenquelle möglich ist. Es ist somit wichtig, dass Metadaten über alle Analyseschritte hinweg verfolgt und gegebenenfalls wieder bereitgestellt werden können.

Im Rahmen eines ODaaS-Systems ist ausschließlich die Verarbeitung der Metadaten notwendig, die für eine weiterführende Bildung von Indikatoren oder für die Analyse von Daten erforderlich sind. Diese Metadaten lassen sich grob in fünf Gruppen unterteilen:

- **Datenquelle:** Diese Art von Metadaten erlaubt es zu identifizieren, wo die Daten ursprünglich erhoben wurden und wer sie aktuell bereitstellt. Besonders wichtig ist dies in Fällen, bei denen Quelle und Bereitsteller sich nicht unmittelbar ergeben, also z.B. bei Derivaten und Analysen mit vererbter Herkunftsinformation. Die Anforderungen, die sich aus der Weiterverarbeitung der Daten ergeben, werden im nachfolgend diskutiert.
- **Inhaltliche Dimension:** Inhaltlich muss zwischen unterschiedlichen Arten von Metadaten unterschieden werden. Dabei handelt es sich zum einen um Metadaten, die Daten kontextualisieren – die also Informationen beinhalten, die sich auf die Bedeutung der Daten beziehen. Sie umfassen beispielsweise Titel und Beschreibung der Daten. Zum anderen handelt es sich um Metadaten, die eine Identifikation der Daten nach Schlagworten ermöglichen. Sie umfassen beispielsweise Klassifikationssysteme oder freie Stichworte.
- **Räumliche Dimension:** Die räumliche Dimension bedingt den Geltungsbereich der Daten. Über Geoinformationen kann beispielsweise der Geltungsbereich auf einzelne Länder oder auch feinere Granularitäten wie Kommunen oder Wahlbezirke eingeschränkt werden.

- **Zeitliche Dimension:** Im Rahmen der zeitlichen Dimension sind zwei Zeitpunkte relevant, die Datenerhebung selbst und der Gültigkeitszeitraum der Daten.
- **Lizenz der Daten:** Dieser Aspekt bezieht sich auf die mit den Daten verbundenen Nutzungsrechte, d.h. auf Einschränkungen oder Erlaubnisse hinsichtlich der weiteren Verwertung der Daten.

Es muss beachtet werden, dass bereits im Revise-Schritt zusätzliche Anforderungen an Metadaten gestellt werden können, um eine automatisierte Verarbeitung der Daten zu ermöglichen. Dabei können sich die Metadatatypen auf verschiedene Dinge beziehen:

- a) auf den gesamten Datenkatalog, also die Gesamtheit aller in einer Open Data-Plattform referenzierten Datenquellen;
- b) auf den Inhalt einer einzelnen Datenquelle, z.B. in Form einer Auflistung der einzelnen hinterlegten Variablen eines Datensatzes und deren Kodierung;
- c) auf einen Datenvektor innerhalb eines Datensatzes, z.B. die Ausgaben unterschiedlicher Resorts einer öffentlichen Verwaltung für ein gegebenes Jahr;
- d) auf einen einzelnen Datenpunkt innerhalb eines Datenvektors.

Ein standardisiertes Metadatenschema für die Beschreibung der Abstraktionsebenen von Datenkatalog und Datensatz wird beispielsweise in der in Kapitel 2 erwähnten W3C-Spezifikation DCAT vorgeschlagen²⁶. Dies gilt jedoch nicht für die Abstraktionsebenen eines einzelnen Datenvektors oder eines einzelnen Datums innerhalb eines Datenvektors.

Werden z.B. Daten einer amtlichen Statistik genutzt und beinhalten diese vorläufigen Prognosewerte, so sind Metadaten auf Ebene der Datenpunkte notwendig. Gleiches gilt, wenn der jeweilige Datensatz Werte aus Imputationsverfahren²⁷ enthält, d.h. Datenlücken durch statistische Verfahren geschlossen wurden. Auch hier handelt es sich um geschätzte Werte, die in nachgelagerten Analysen ggf. anders verarbeitet werden müssen.

Je nach Analysezusammenhang und konkreter Ausgestaltung von ODAaaS kann es notwendig sein, dass Metadaten in hoher Granularität, ggf. mit Bezug auf einzelne Datenpunkte, vorliegen müssen. Die Qualitätsanforderungen an ein Analyseergebnis müssen im Einzelfall bewertet werden.

ODAaaS-Konzepte, die ein hohes Wertschöpfungspotential offener Daten realisieren sollen, müssen die im Retrieve-Schritt verwendeten Metadaten im Revise-Schritt um entsprechende Metadaten auf allen Ebenen, also Datenkatalog, Datensatz, Datenvektor und Datum, anreichern.

Die Anforderung der Zusammenstellung feingranularer Metadaten zu einem Datensatz wird umso relevanter, je höher die Ansprüche an komplexe analytische Verfahren und an die Zusammenführung von Daten im Rahmen einer konkreten Ausgestaltung von ODAaaS sind. Denkbar sind organisatorische Ansätze, die die Integration einer Community von Datennutzern in den Prozess des Aufbaus eines ODAaaS-Ökosystems beinhalten. Alternativ sind technische Ansätze denkbar, die teilautomatisiert zusätzliche Metadaten aus Daten extrahieren. Grundsätzlich muss in diesen Prozess umfangreiches Domänenwissen einbezogen werden.

Datenqualität

Ein Mindestmaß an Datenqualität ist für diesen sowie alle nachgelagerten Schritte eine notwendige Anforderung. Auf Open Data-Plattformen werden oft Daten aus unterschiedlichen Quellen zusammengeführt. Der Aufwand für die Sicherstellung der Datenqualität steigt mit der Anzahl unterschiedlicher Datenquellen.

²⁶ W3C Recommendation „Data Catalog Vocabulary“ (DCAT), vgl. <http://www.w3.org/TR/vocab-dcat/>.

²⁷ Vervollständigung unvollständiger Datensätze mittels Schätzwerten oder Simulationen

Wenn in diesem Kontext von Potentialen und Anforderungen hinsichtlich der Datenqualität gesprochen wird, so muss auf unterschiedliche Dimensionen von Datenqualität eingegangen werden. Die Ursachen für Schwächen im Bereich der Datenqualität liegen oft darin begründet, dass

- unterschiedliche Anforderungen an die Maschinenlesbarkeit und automatisierte Verarbeitbarkeit existieren;
- im Integrate-Schritt Daten aus unterschiedlichen Quellen zusammengeführt werden;
- bei den Bereitstellern offener Daten die erforderlichen Kompetenzen für Datenhaltung und professionelles Datenmanagement nicht immer gegeben sind.

Was die Datenqualität im Sinne einer **inhaltlichen Korrektheit** und Fehlerfreiheit in Bezug auf falsch übertragene oder falsch erhobene Werte angeht, sind sowohl technische als auch organisatorische Anforderungen beim Datenbereinsteller zu sehen. Die organisatorischen und technischen Anforderungen stehen dabei in einem sich wechselseitig bedingenden Verhältnis: je stärker der Prozess der Bereitstellung auf Basis technischer Systeme realisiert wird, desto eher muss der Fokus auf die Infrastruktur zur Bereitstellung der Daten gelenkt werden. Ist die Bereitstellung weniger durch technische Mittel hinterlegt, muss organisatorisch durch klare Verhaltensregeln die Qualität der bereitgestellten Daten gewährleistet werden.

Ähnliche Voraussetzungen bezüglich der Zuständigkeit für die Sicherstellung einer hinreichenden Datenqualität bestehen für die **Vollständigkeit** der Daten. Auch hier liegen die organisatorischen und technischen Anforderungen auf Seiten des Datenbereitstellers. Dabei kommen z.B. bei der amtlichen Statistik Regularien zum Tragen, die eine Veröffentlichung von Daten nur dann zulassen, wenn gewisse Mindestkriterien, z.B. in Bezug auf Repräsentativität, erfüllt sind. Daten können unzulässig sein, auch wenn sie aus Sicht des Datenbereitstellers unter Berücksichtigung anderer Kriterien als vollständig erachtet werden. Dies gilt jedoch nur für einen eher eingeschränkten Kreis der veröffentlichten Datenbestände, so dass hier – die Erfüllung aller weiteren Qualitätsmerkmale vorausgesetzt – keine zusätzlichen spezifischen Anforderungen für ODAAaaS entstehen. Generell gilt, dass mit der Vollständigkeit von Daten deren Nutzbarkeit steigt, d.h. je präziser und vollständiger die zur Verfügung gestellten Informationen vorliegen, desto eher können Informationen in Analysesystemen genutzt werden. Andererseits existieren spezielle Analyseverfahren die insbesondere zur Auswertung unvollständiger und widersprüchlicher Daten entwickelt worden sind.

Das Untergliederungsmaß von Datenstrukturen von bereitgestellten offenen Daten wird durch die **Granularität** beschrieben. Für georeferenzierte Daten spielt nicht nur ihre Genauigkeit sondern auch die Granularität eine wichtige Rolle. Feingranulare Adressangaben wie Straße und Hausnummer werden in höheren Aggregationsstufen ggf. nur noch als Städte oder Bundesländer referenziert, wobei in letzterem Fall unterschiedliche Positionen herangezogen werden können, um die Daten geographisch beschreibbar zu machen.²⁸

Ein weiterer wichtigerer Aspekt der Datenqualität betrifft die **Wohlgeformtheit**²⁹ der Daten. Eine besondere Anforderung ergibt sich, wenn Daten erst durch den Bereitsteller in maschinenlesbare Form überführt werden. Die größten Probleme bezüglich der Wohlgeformtheit der Daten entstehen insbesondere durch Fehler, die sich bei der Konvertierung der Daten in veröffentlichungskonforme Formate ergeben. Die Wohlgeformtheit der Daten ist ein notwendiges, aber kein hinreichendes Kriterium für die automatisierte Maschinenlesbarkeit von Daten. Es können insbesondere bei der Konvertierung zwischen Datenformaten Fehler entstehen, die sich negativ auf die Integrität von Daten auswirken.

²⁸ In der Praxis werden hierbei auf höheren Ebenen „Zentroide“ herangezogen. Darunter versteht man den Schwerpunkt einer räumlich oder durch einen Polygonzug abgegrenzten Fläche. Bei der Betrachtung von Bundesländern können andere relevante Positionen wie die Landeshauptstadt herangezogen werden.

²⁹ Wohlgeformte Daten genügen speziellen syntaktischen Regeln.

Die Wohlgeformtheit von Datensätzen sollte bereits im Rahmen der Datenbereitstellung kontrolliert werden. Daten, die diesbezüglich Schwächen aufweisen, sind in nachfolgenden Schritten schwerer zu verarbeiten.

In einem ODAaaS-Gesamtsystem müssen für den Revise-Schritt Werkzeuge zur Adressierung der genannten technischen Anforderungen vorgehalten werden. Im Kontext offener Daten sind dabei mehrere Varianten denkbar, z.B. die Möglichkeit, Lösungen zur Sicherung der Datenqualität direkt in der Daten bereitstellenden Organisation zu betreiben. Alternativ können entsprechende Qualitätssicherungslösungen von Drittanbietern im Revise-Schritt „as a Service“ in Anspruch genommen werden.

Alternativ können Speziallösungen der Datenbereinigung im Revise-Schritt direkt angewendet werden. Hier obliegt es dem Anwender, eventuelle Probleme der Datenqualität zu erkennen und mit Hilfe der zur Verfügung stehenden Lösungen zu bearbeiten. „Cleaning-Lösungen“ liegen bereits bei nahezu allen kommerziellen Anbietern als Bestandteil von Datenanalyselösungen vor. Sie erfordern jedoch meist Spezialkenntnisse, was wiederum den Kreis der Nutzer einschränkt.

Datenprovenienz

Die Provenienz von Daten, d.h. die über unterschiedliche Prozessschritte hinweg mögliche Zuordnung der Datenherkunft und damit verbundener Informationen, ist eine wesentliche Informationsquelle. Sie wird z.B. in der W3C DCAT-Spezifikation mit Hilfe von Metadaten nachvollziehbar gemacht. Vor dem Hintergrund der Zusammenführung von Daten bieten die Analyseschritte einen hohen Wert. In einem ODAaaS-System sollte daher bei Transformation und der Verarbeitung von Daten im Rahmen der Analyse die Möglichkeit bestehen, dass entsprechende Herkunftsinformationen weitergeleitet und den Derivaten aus der Zusammenführung von Daten zugeordnet werden können. So stehen diese Informationen den Anwendern auch im Rahmen der Interpretation von Analyseergebnissen zur Verfügung. Da im Revise-Schritt nur Daten aus einer Quelle bearbeitet werden existieren bezüglich der Datenprovenienz keine besonderen Anforderungen. Ein im Folgeschritt „Integrate“ zusammengeführter Datensatz muss jedoch über mehrere Provenienzinformationen verfügen können.

Zusammenfassung

Zusammengefasst lassen sich für den Revise-Schritt folgende Aussagen treffen:

- Daten werden auf bestimmte Aspekte der Datenqualität geprüft, d.h. für spätere Schritte werden möglichst widerspruchsfreie Daten bereitgestellt.
- Metadaten werden auf Vollständigkeit geprüft und im Falle komplexer Anwendungsszenarien um Metadaten auf der Ebene einzelner Datenvektoren und Datenpunkte ergänzt.
- Es wird sichergestellt, dass Metadaten über mehrere Prozessschritte hinweg angereichert werden können, ursprüngliche Metadaten jedoch erhalten bleiben.

4.3 Datenzusammenführung - Integrate

Liegen offene Daten in einem ODAaaS-Ökosystem integriert vor, können höhere Nutzungspotentiale realisiert werden. In diesem Fall besteht die Möglichkeit, Analysen nicht nur auf einen einzelnen Datensatz, sondern auf zusammengeführte Datensätze durchzuführen. Hierbei ist die Integration nicht auf offene Daten beschränkt. Funktionalitäten von ODAaaS-Systemen können auch darin bestehen, offene Daten und nicht-offene Daten zu integrieren. Je nach Anwendungsfall können unterschiedliche Grundanforderungen für den Integrationsschritt vorliegen:

- Es werden nur offene Daten aus einer einzelnen Datenquelle (im Sinne einer Datenressource) in einer Open Data-Plattform referenziert. In diesem Fall findet keine Integration statt. Dies stellt jedoch einen in der Praxis selten vorkommenden Sonderfall dar.
- Es werden offene Daten aus mehreren Datenquellen in einer Open Data-Plattform referenziert. Die einzelnen Datensätze werden jedoch nicht gemeinsam analysiert. Analysen beziehen sich jeweils nur auf einen einzelnen Datensatz. Auch in diesem Fall besteht keine Notwendigkeit der Integration der Daten.
- Es werden offene Daten aus mehreren Datenquellen referenziert. Analysen beziehen sich auf mehrere Datensätze, die untereinander verbunden werden müssen. In diesem Fall besteht die Notwendigkeit der Integration der offenen Daten.
- Es werden offene Daten aus mehreren Quellen referenziert und mit nicht-offenen, privaten Daten zusammengeführt. Analysen beziehen sich auf die Zusammenführung aus offenen und nicht-offenen Daten. In diesem Fall besteht sowohl die Notwendigkeit der Integration offener Daten untereinander als auch die zusätzliche Integration der nicht-offenen, privaten Daten.

Zentrale und dezentrale Integration von Daten

Die Integration von Daten kann prinzipiell auf zwei Arten erfolgen: zentral oder dezentral. Bei zentralen Ansätzen werden Daten aus unterschiedlichen Quellen zusammengeführt und in einer zentralen Infrastruktur, z.B. einem Data Warehouse zusammengeführt und hinterlegt. Bei einer dezentralen Integration von Daten verbleiben die Daten in den jeweiligen Quellen. Es findet keine Hinterlegung der Daten in einer zentralen Infrastruktur statt. Die Daten werden dort nur referenziert. Dabei hat es für den Nutzer den Anschein, als wären die Daten tatsächlich in einer zentralen Infrastruktur hinterlegt.

Unabhängig vom jeweils gewählten Ansatz kann der Aufwand der Integration durch die Verwendung von Linked Open Data reduziert werden. Inhaltliche Querbezüge zwischen den einzelnen Datensätzen sind hierbei bereits vorhanden. Relationen zwischen zu integrierenden Daten lassen sich leicht herstellen. Der Anteil an offenen Daten, die als Linked Open Data vorliegen, ist bisher jedoch vergleichsweise gering.

Ein möglicher Ansatz zur zentralen Integration von Daten besteht im Einsatz von ETL-Lösungen. ETL-Lösungen erlauben es, Daten aus unterschiedlichen Datenbanken zu extrahieren (englisch: extract), in ein entsprechendes Zielformat zu übertragen (englisch: transform) und diese in einen Zieldatenspeicher zu übertragen (englisch: load)³⁰. Als Ergebnis liegt dann eine Zieldatenbank vor, welche die Daten für den Nutzer zugänglich macht. ETL-Lösungen werden hauptsächlich für den Aufbau von Data Warehouse-Architekturen eingesetzt. Unter einem Data Warehouse ist in diesem Kontext eine integrierte Dateninfrastruktur zu verstehen, welche sich aus unterschiedlichen Datenquellen zusammengeführt ist. Innerhalb eines ODAAAS-Ökosystems können im Integrate-Schritt in regelmäßigen Abständen die Metadatenkataloge von Open Data-Plattformen ausgelesen werden und die Daten durch ETL-Prozesse in eine zentrale Dateninfrastruktur überführt werden.

ETL-Lösungen können bereits im Revise-Schritt im Rahmen Qualitätsprüfungen und Transformationen vornehmen und sind bei vielen kommerziellen Lösungen auch Bestandteil des Funktionsumfangs. Die wesentliche Eigenschaft von ETL-Lösungen besteht darin, dass Daten unabhängig von der Verfügbarkeit von Quelldaten verwendet werden können. Werden auf Seiten der Datenbereitstellung Veränderungen vorgenommen stehen die Daten, die im Rahmen eines ETL-Prozesses in ein Zielsystem übertragen wurden, weiterhin in unveränderter Form zur Verfügung. ETL-Lösungen lassen sich gut in Cloud-Infrastrukturen integrieren. Der Aufwand bei ETL-Lösungen steigt proportional zu der Anzahl der zu integrierenden Datensätze. Für jeden Datensatz sind zugehörige

³⁰ In unserer Terminologie entspricht das den Schritten Retrieve – Revise – Integrate. Im Rahmen der Verarbeitung großer Datenmengen haben sich ergänzend ELT-Verfahren etabliert, die die vorbereitenden Schritte für eine Datenanalyse in der Reihenfolge Retrieve – Integrate – Revise durchführen. Dieser Ansatz ist zumeist mit der Nutzung extrem schneller in-memory-Datenbanken verbunden.

Transformationsregeln zu implementieren und bei der Vernetzung dieser Daten entsprechende Relationen in der Zieldatenbank einzuarbeiten. Eine eher negative Eigenschaft von ETL-Lösungen ist, dass im Falle der Integration großer Datenbestände viele Daten zwischen Datenquellen bewegt werden müssen, unabhängig vom Umfang der späteren Nutzung.

Data Federation-Konzepte stellen eine dezentrale Lösung dar. Daten aus unterschiedlichen Quellen werden dabei nicht in eine zentrale Dateninfrastruktur überführt sondern in einheitlicher Weise in einer virtualisierten Form zur Verfügung gestellt. Virtualisiert bedeutet in diesem Kontext, dass der Nutzer keine Information darüber benötigt, wo die Daten konkret hinterlegt sind und in welchen Formaten sie vorliegen. „Data Federation“ stellt eine Abstraktionsebene oberhalb der Datenquellen zur Verfügung. Das Ergebnis eines solchen Ansatzes wird als Federated Database System (FDBS) bezeichnet. Da bei einem FDBS keine explizite Speicherung der zusammengeführten Daten erfolgt, ist die Nutzung eines Storage as a Service-Diensts nicht erforderlich. Die Daten werden im nachfolgenden Analyse-Schritt Ad-hoc übertragen und verarbeitet. In einer FDBS-Lösung werden Daten somit direkt aus den verschiedenen Quellen bezogen und dem Nutzer zur Verfügung gestellt. Die in den Quellen vorhandenen Metadaten müssen ergänzend übertragen und im FDBS hinterlegt werden.

Data Federation-Ansätze können klar von ETL-Ansätzen abgegrenzt werden. ETL-Lösungen werden dazu genutzt, eine Zieldatenbank, wie z.B. ein Data Warehouse, zu erstellen. Im Gegensatz dazu werden mit Data Federation-Ansätzen keine umfassenden Zieldatenbanken erstellt. Die für eine weitere Verarbeitung benötigten Daten werden stattdessen im Analyse-Schritt aus den ursprünglichen Quellen d.h. dem Ergebnis des Revise-Schrittes, bezogen. Der wesentliche Vorteil von Data Federation-Lösungen gegenüber ETL-basierten Lösungen liegt somit darin, dass selten oder nie analysierte Daten nicht auf Verdacht in eine Zieldatenbank transferiert werden müssen. Dies spielt insbesondere dann eine Rolle, wenn es sich um große Datenbestände handelt. Der Nachteil von Data Federation-Lösungen besteht darin, dass bei unterschiedlichen Analyseschritten der Integrationsvorgang und die damit verbundenen Anpassungen der Daten erneut durchgeführt werden müssen. Während bei ETL-Lösungen Änderungen der primären Quelldaten bei der Analyse nicht bemerkt werden, machen sich Änderungen bei FDBS-Lösungen in der Analyse bemerkbar.

Bei der Etablierung eines ODAaaS-Ökosystems sind Maßnahmen für die Bereitstellung qualitativ hochwertiger Daten, die bereits möglichst einheitlichen Datenmodellen und Metadatenformaten genügen, geeignet, um den Integrations- und Anpassungsaufwand zu reduzieren. Wie bereits im Revise-Schritt ist auch für die widerspruchsfreie Integration und Verknüpfung von Daten ein erhebliches Maß an Domänenwissen notwendig.

Integration als Dienst

Die Datenintegration kann im Rahmen von Community-Aktivitäten verteilt durchgeführt werden. Weiterhin macht es Sinn, das Ergebnis des Integrationsschrittes verschiedenen Nutzergruppen für deren Analyse-Schritte zur Verfügung zu stellen. Nutzer können Daten so für spezifische Fragestellungen aufbereiten (Revise, Integrate) und als Information as a Service zur Verfügung stellen. Sofern die Aufbereitung automatisierbar ist, kann eine Information as a Service-Plattform als Laufzeitumgebung für Analyseverfahren bereitgestellt werden. In Fällen, in denen ein erhöhtes Maß an manueller Aufbereitung von Daten notwendig ist, sollte den Nutzern eine Umgebung zur Aufbereitung sowie die aufbereiteten Daten als Information as a Service-Datendienst zur Verfügung gestellt werden. Bei diesem Ansatz muss jedoch die Nachvollziehbarkeit der Datenaufbereitung über Metadaten gewährleistet sein.

Eine weitere Variante der Einbeziehung nutzergenerierte Inhalte ist in der Nutzung von Ontologien. Sie ermöglichen die Transformation von Variablen und vereinfachen die Zusammenführung von Daten.

Zusammenfassung

Zusammengefasst lassen sich für den Integrate-Schritt folgende Aussagen treffen:

- Bei Verwendung von ETL-Ansätzen liegen Daten nach Retrieve-, Revise- und Integrate-Schritten an einem zentralen Speicherort vor. Die Aktualität der Daten ist nicht automatisch gewährleistet.
- Bei Verwendung von FDBS-Ansätzen liegen die Daten verteilt vor. Die Aktualität der Daten ist gewährleistet, dafür sind jedoch die Retrieve-, Revise- und Integrate-Schritte bei jeder Analyse erneut zu durchlaufen.
- Daten können durch Nutzer oder Communities verändert, transformiert und angereichert werden.

4.4 Datenanalyse - Analyze

Im Analyse-Schritt stehen nicht einzelne Methoden oder statistische Verfahren im Vordergrund, sondern der Analyseprozess im Ganzen. ODAAaaS-Systeme unterscheiden sich in der Art und Weise, wie der Analyseprozess organisatorisch durchgeführt wird und welche Akteure in welcher Weise beteiligt sind.

Eigenschaften generischer und spezifischer Analysewerkzeuge

Bei der Durchführung von Datenanalysen stellt sich die Frage, ob allgemeine, generische oder fachlich spezifische Analysewerkzeuge genutzt werden sollen. Die Entscheidung hat konkrete Auswirkungen auf die Ausgestaltungsformen von ODAAaaS.

- Generische ODAAaaS-Lösungen zielen überwiegend auf die vollständige Analysierbarkeit offener und privater Daten und die Visualisierbarkeit der Analyseergebnisse ab.
- Spezifische ODAAaaS-Lösungen beziehen sich auf einen klar abgrenzbaren Teil offener und privater Daten und stellen auf die spezifische Fachdomäne zugeschnittene Analyse- und Visualisierungswerkzeuge zur Verfügung.

Die Entscheidung für eine der beiden Ausgestaltungsformen hängt stark von den Anforderungen des Nutzers bzw. dem Geschäftsmodell des Anbieters ab. Die technischen und inhaltlichen Anforderungen an die vorgelagerten Revise- und Integrate-Schritte sind für spezifische ODAAaaS-Lösungen geringer, da sie sich nur auf eine wohldefinierte und im Umfang beschränkte Teilmenge offener Daten beziehen. Anwendbarkeit und Nutzen spezifischer Lösungen sind jedoch durch die Fokussierung auf eine Fachdomäne beschränkt. Der Analyse-Schritt kann eine hohe fachliche Komplexität aufweisen, je nachdem welche Analysemethoden angewendet werden bzw. in welcher Art die Ergebnisse zur Verfügung gestellt werden. Die Bereitstellung spezifischer ODAAaaS-Lösungen als Cloud-Dienste stellt bei einer ausreichenden Anzahl potentieller Nutzer eine sinnvolle technische und wirtschaftliche Alternative dar.

Generische ODAAaaS-Lösungen können allgemein zur Analyse offener Daten eingesetzt werden und adressieren grundsätzlich die Gesamtheit aller Daten. Sie stellen höhere Anforderungen an die alle Prozessschritte, da keine Annahmen über die Art der zu analysierenden Daten und die Form der Analyse gemacht werden können. Cloud-Lösungen zur Bereitstellung und Speicherung der Daten sowie zur Bereitstellung von Analysemethoden sind für diese Ausgestaltungsformen gut geeignet.

Nutzerspezifische Ausgestaltung von Analysemethoden

Ein grundlegendes Ausgestaltungsmerkmal des Analyse-Schritts besteht im Umfang der angebotenen Analysemethoden bzw. in der Möglichkeit, Analysefunktionen selbst zu implementieren. Während Nutzer in einem geschlossenen Analysesystem wenig Einfluss auf den Funktionsumfang haben, können Nutzer in offenen Systemen Analysefunktionen aktiv bereitstellen. Geschlossene ODAAaaS-Systeme können als Spezialfall von SaaS angeboten werden. Offene ODAAaaS-Systeme besitzen zusätzlich PaaS-Eigenschaften. Entsprechende Konzepte werden in statistischen Umgebungen bereits erfolgreich umgesetzt.

Im Fall offener Systeme können von Nutzern bereitgestellte Funktionen auch anderen Nutzern als SaaS-Dienste zur Verfügung gestellt werden. Wie in geschlossenen Systemen müssen auch alle in offenen Systemen angebotenen Analysefunktionen über eine fachgerechte und umfassende Nutzerdokumentation verfügen.

Insbesondere generische Analysesysteme können auf Funktionen beschränkt sein, die durch den Nutzer nicht kombiniert eingesetzt werden können. Sie dienen meist einer Beschreibung einfacher Zentralwerte oder Verteilungsparameter von Daten. Im Allgemeinen sind sie auf einzelne Datensätze beschränkt, erlauben aber im Einzelfall das Filtern von Daten nach vorgegebenen Kriterien. Der Eingriff des Nutzers beschränkt sich in der Regel auf die Auswahl der darzustellenden Variablen, die Auswahl von Filterkriterien und gegebenenfalls die Auswahl der Darstellungsform. Derartige Systeme geben dem Nutzer einen relativ geringen Spielraum, um die Daten für weitergehende Analysen zu nutzen. Sie haben meist den Charakter von einfachen Informationssystemen, können jedoch bei Bereitstellung geeigneter APIs als Information as a Service-Plattformen angeboten werden.

Geschlossene und insbesondere spezifische ODAaaS-Systemen sind stark an einen bestimmten Anwendungsfall gebunden und weisen wenig Flexibilität auf. Sie sind vor allem für Fachnutzer mit geringer statistischer Kompetenz geeignet. Fachliche Anforderungen bestehen in der Festlegung sinnvoll anzuwendender Modelle, Kennzahlen und Visualisierungsformen. Bei fehlenden Metadaten können Filterung, Analyse und Visualisierung nur mittels generischer Operationen durchgeführt werden.

Wird ein ODAaaS-System als modulare PaaS-Umgebung angeboten, stehen dem Nutzer kombinierbare Analyse- und Visualisierungsdienste zur Verfügung. Die kombinierten Dienste können, ggf. ergänzt um nutzerspezifische Methoden, als SaaS genutzt werden. Die Benutzeroberfläche kommuniziert, wie bei SaaS-Diensten üblich, mittels Web-Technologien mit den angebotenen Diensten. Je nach genutztem grafischen Framework und angebotenen Schnittstellen können Teile der Visualisierung auch nutzerseitig als Funktionalität des Klienten im Browser angeboten werden. Modulare PaaS-Ansätze bieten eine hohe Flexibilität bei Auswertung und Darstellung von Ergebnissen. Sie erlauben dem Nutzer, eigene Analysemethoden zu implementieren und spezifische Analysen durchzuführen. Dies erhöht die Freiheitsgrade des Nutzers und ist insbesondere für Nutzer mit technischer und fachlicher Kompetenz interessant.

Datenmarktplätze

Wird ein ODAaaS-System in Form einer offenen, verteilten Dienstarchitektur realisiert, in der Dienste durch Schnittstellen frei kombinierbar sind, so stehen Diensteanbietern und Dienstonutzern zahlreiche Freiheitsgrade zur Verfügung, um die Dienste zu implementieren, zu Softwareangeboten zu bündeln und auf offene Daten anzuwenden. Diese Ausgestaltungsform von ODAaaS kann als dezentral organisierte, cloudbasierte Infrastruktur im Sinne eines Daten- und Dienstmarktplatzes angesehen werden. Dabei werden veredelte und integrierte Daten mittels Information as a Service-Diensten zur Verfügung gestellt. Angebotene Filter-, Analyse- und Visualisierungsdienste können in einer PaaS-Umgebung durch den Nutzer kombiniert und durch eigene Dienste ergänzt werden. Der Endnutzer führt die Kombination der Dienste zumeist nicht selber durch. Stattdessen greift er auf Angebote zurück, die durch technisch orientierte Nutzer auf Basis bestehender Dienste erstellt werden. Dieses erlaubt eine hohe Flexibilität, erfordert jedoch das Vorhandensein abgestimmter technischer Schnittstellen zwischen allen Teilnehmern auf dem Marktplatz. Der konzeptionell vorhandenen Flexibilität einer derartigen Lösung steht die hohe technische Komplexität gegenüber. Die Realisierung eines verteilten, föderierten ODAaaS-Marktplatzes ist daher unwahrscheinlich.

Speicherung von Ergebnisdaten

Die in den Schritten *Revise* und *Integrate* erzeugten Daten können in Datenbanken oder mittels *Storage as a Service*-Diensten persistiert werden. Gleiches gilt auch für im *Analyse*-Schritt durch Filterung oder Transformationen erzeugte Zwischenergebnisse. So ist es möglich, dass integrierte Daten voranalysiert und

innerhalb von ODAaaS für nachfolgende Analysen persistiert werden oder im Sinne eines Datenmarktplatzes über spezielle APIs anderen Nutzern oder Analysewerkzeugen als Information as a Service zur Verfügung stehen.

Bei Lösungen, die nur einfache Ad-hoc-Analysen offener Datensätze erlauben, ist es im Allgemeinen nicht notwendig, dass Zwischenergebnisse im System vorgehalten werden. Steigen jedoch die Komplexität der Analysen, die Menge der vorausgewerteten Daten und insbesondere der Umfang der einbezogenen privaten Daten, so wird eine persistente Speicherung der Zwischenergebnisse sinnvoll. Wird ein derartiges Analysesystem als SaaS angeboten, so ist seine Mandantenfähigkeit zu garantieren. Während die integrierten offenen Daten noch von verschiedenen Nutzern analysiert werden können, stehen die Zwischenergebnisse zunächst nur deren Erzeuger zur Verfügung. Dies gilt insbesondere bei der Einbeziehung privater Daten, die jedoch bereits die allgemeine Nutzung der im Integrate-Schritt zusammengeführten Daten einschränkt.

Zusammenarbeit und Community-Ansätze

Bei der gemeinsamen Auswertung offener Daten durch mehrere unabhängige Personen oder bei der gemeinsamen Auswertung offener und privater Daten durch mehrere Mitarbeiter einer Einrichtung treten neue Fragestellungen bezüglich der Ausgestaltung von ODAaaS auf.

Fach- und Domänenwissen sind für avancierte Ausgestaltungsformen der Werkzeuge des Analyse-Schritts unumgänglich. Ist der Grad des notwendigen Domänen- und Fachwissens zur Durchführung von Analysen hoch, werden nur kleinere Zielgruppen angesprochen was eine wirtschaftliche Anforderung für den Anbieter der Werkzeuge darstellt. Weiterhin können sowohl technische als auch organisatorische Anforderungen abgeleitet werden, je nachdem, in welchem Umfang die Nutzer offene ODAaaS-Systeme durch eigene Werkzeuge anreichern können oder bei der Datenanalyse zusammenarbeiten. Dabei macht es einen Unterschied, ob Nutzer nur mit den Anbietern der Werkzeuge kommunizieren können oder bei der Analyse interaktiv in Communities zusammenarbeiten können.

Die Zusammenarbeit innerhalb einer Community ist auf mehrere Arten vorstellbar. Im einfachsten Fall stellt die Community einen übergreifenden Wissensträger dar. Dies kann durch die Einbeziehung einfacher Kollaborationswerkzeuge wie Wikis in ODAaaS gewährleistet werden. Weiterhin kann das Wissen zu unterschiedlichen Analysemethoden oder fachspezifischen Vokabularen im Sinne von Best-Practice-Anleitungen, Wissensbasen, Taxonomien und Ähnlichem zur Verfügung gestellt werden. Technische Anforderungen ergeben sich aus Bereitstellung zugehöriger Bausteine, organisatorische Anforderungen aus der Pflege der Inhalte und der Sicherstellung von Korrektheit und Konsistenz. In geschlossenen ODAaaS-Systemen entfallen alle mit dem Einbringen eigener Werkzeuge verbundenen technischen und organisatorischen Anforderungen.

Gemeinsames Arbeiten, d.h. die Zusammenarbeit mehrerer Nutzer im Revise-Schritt (Daten-Cleaning, Verbesserung der Metadaten), im Integrate-Schritt (Zusammenführung von Daten, Erstellung von Metadaten) und im Analyse-Schritt (Aufbereitung der Daten für spezifische Fragestellungen, Generierung einer Analysestrategie, Interpretation der Ergebnisse) bedingt die Bereitstellung zugehöriger Lösungen in ODAaaS. Dabei ist zu unterscheiden, ob die Zusammenarbeit in einen geschlossenen Nutzerkreis, wie einer Gruppe von Mitarbeitern einer Einrichtung, oder in einem offenen Nutzerkreis, wie einer über eine gemeinsame Zielsetzung definierten Community, erfolgen soll. In Abhängigkeit von den Anforderungen des Nutzerkreises muss ein ODAaaS-System entsprechende Rollen- und Berechtigungskonzepte anbieten.

4.5 Akteure im Umfeld von ODAAaaS

Eine detaillierte Analyse der in ODAAaaS vorhandenen Rollen und deren inhaltliche, organisatorische und wirtschaftliche Zusammenhänge setzt eine präzise Definition der beteiligten Akteure voraus. Im Folgenden wird eine an die beschriebenen Prozessschritte und an Dienstmodelle im Cloud-Computing angelehnte Definition gegeben.

Dienstnutzer

Dienstnutzer führen einen oder mehrere Prozessschritte in ODAAaaS aus und greifen dabei auf die Ergebnisse des vorhergehenden Analyseschritts zurück. Für die Ausführung ihres Schrittes nehmen sie zugehörige von ODAAaaS bereitgestellte Dienste in Anspruch. Jeder Dienstnutzer kann prinzipiell durch das Angebot ermittelter Zwischenergebnisse oder die Weitergabe eigener Analysewerkzeuge zum Dienstanbieter werden. Die nachfolgend aufgelisteten Dienstnutzer verwenden demzufolge einen Analysedienst zur Ausführung ihres spezifischen Prozessschritts und nehmen ggf. ergänzend einen Cloud-Dienst zur Speicherung ihrer Zwischenergebnisse in Anspruch.

- Datenkonsument (englisch: open data consumer, data storage consumer)
- Datenbereiniger (englisch: revision service consumer, data storage consumer)
- Datenintegrator (englisch: integration service consumer, data storage consumer)
- Datenanalyst (englisch: analytics service consumer)

Dienstanbieter

Dienstanbieter stellen zum einen Rohdaten oder veredelte Daten bzw. Informationen zur Verfügung. Zum anderen bieten sie Cloud-Dienste zur Durchführung der einzelnen Analyseschritte an.

- Datenbereitsteller (englisch: open data provider, data storage consumer)
- Informationsbereitsteller (englisch: information provider)
- Anbieter von Datenspeicher (englisch: data storage provider)
- Anbieter von Datenveredlungsdiensten (englisch: revision service provider)
- Anbieter von Datenintegrationswerkzeugen (englisch: integration service provider)
- Anbieter von Analysewerkzeugen (englisch: analytics service provider)

Im Sinne der Fragestellung des Reports geht es primär um Vorteile für Datenanalysten in der Rolle des „Analytics Service Consumer“ und um Betriebs- und Geschäftsmodelle für „Analytics Service Provider“. Die weiteren Rollen ermöglichen eine Detaillierung dieses Rollenkonzepts in Bezug auf die Prozessschritte bei der Analyse offener Daten.

Vernetzte Wertschöpfungsketten

Wie vorgehend dargestellt, können die einzelnen Prozessschritte in ODAAaaS im einen Extrem durch einen Endnutzer, der ein homogenes Werkzeug eines Anbieters nutzt, durchgeführt werden. Im anderen Extrem können die einzelnen Schritte von einer Nutzer-Community durchgeführt werden, die föderierte Werkzeuge bzw. Cloud-Dienste verschiedener Anbieter nutzt.

Für föderierte Systeme ergibt sich die Frage, in welcher Weise die unterschiedlichen Dienste untereinander bzw. gegenüber dem Endkunden eingepreist werden können. Hierdurch ergeben sich außerdem ausgeprägte Querbezüge zu technischen und organisatorischen Aspekten und einer damit einhergehende höhere Anforderung an Harmonisierung und Interoperabilität der einzelnen Dienste zueinander. Wertschöpfungsketten ergeben sich dabei zwischen den beteiligten Akteuren wie den unterschiedlichen Dienst Anbietern, Dienst-Aggregatoren und Dienstnutzern.

Abrechnungsmodelle

Im Falle der Kommerzialisierung von ODAaaS Lösungen spielt die Möglichkeit, geeignete Abrechnungsmodelle zu implementieren, eine wichtige Rolle. Hierbei ist zu berücksichtigen in welcher Weise ein ODAaaS-System ausgestaltet ist. Bei föderierten Ansätzen sind gegenseitige Abrechnungsmöglichkeiten vorzusehen, sofern dies lizenzrechtlich und wirtschaftlich möglich und sinnvoll ist. Wird ODAaaS als standardkonformes, föderiertes Cloud-System angeboten, so sollten die dort zwischen Dienst Anbietern und Kunden vorhandenen Abrechnungsmodelle ausreichend sein, um den Anforderungen von ODAaaS zu genügen.

5 Nutzen von Datenanalyse im Bereich Open Data

Ein Ziel der vorliegenden Studie ist es, Potentiale und Hindernisse für Werkzeuge zur Datenanalyse als Dienstleistung aufzuzeigen. Es wird untersucht, wie bei stetig steigenden Datenmengen die Auswertung offener Daten für verwaltungsinterne und –externe Akteure möglichst einfach gestaltet werden kann. Diese Untersuchung wird für die Anbieter und Nutzer der Dienste zur Durchführung der einführend identifizierten Prozessschritte getrennt durchgeführt. Für die vollständige oder teilweise Bereitstellung von ODAAA-S-Komponenten durch Cloud-Dienste im Rahmen einer dienstorientierten Architektur (SOA) ergibt sich für die beteiligten Unternehmen in stärkerem Maße die Möglichkeit ihre Dienste am Markt zu platzieren.

Generell eröffnet die moderne Datenanalyse eine Vielzahl von Möglichkeiten, um Daten nutzbringend einzubringen und auszuwerten. Hierbei können bereits einfache statistische Operationen einen enormen Qualitätsgewinn gegenüber Daten im Rohformat darstellen. Weiterhin steigt der Nutzen bei der Analyse von Daten mit Umfang und Qualität der zugehörigen Metadaten und dem Grad der Zusammenführung von Daten aus unterschiedlichen Quellen.

Im Folgenden wird an ausgewählten Beispielen aufgezeigt, in welchen Kontexten ODAAA eingesetzt werden kann und welcher übergeordnete Nutzen sich durch die Analyse offener Daten ergibt.

5.1 Übergeordnete Nutzenpotentiale

Zunächst soll auf die Frage eingegangen werden, welcher Zugewinn durch die Anwendung der vier Prozessschritte gegenüber der reinen Vorhaltung offener Daten erreicht werden kann.

Anreicherung und Veredelung offener Daten

Im Prozessschritt *Revise* erfolgt eine Anreicherung und Veredelung der offenen Daten. Dieser Schritt ist häufig für Nutzer ohne entsprechende fachliche und technische Kenntnisse nur schwer zu bewältigen. Im Rahmen von ODAAA bieten sich umfangreiche Optionen für die Anreicherung offener Daten an. Zu diesem Zweck existieren am Markt umfangreiche und mächtige Werkzeuge.

Wird ODAAA so eingesetzt, dass Nutzer veredelte Daten im *Revise*-Schritt wieder selbst zur Verfügung stellen können, so wird das Erkenntnisinteresse weiterer Kreise befriedigt. Dies hat wesentliche Vorteile. Durch die Integration der Anwenderperspektive kann die Weiternutzung veredelter Daten klarer und fachlich aufbereiteter erfolgen, als dies bei Rohdaten der Fall ist.

Transformation von Daten zu Information und Wissen

Daten bieten eine gute Ausgangslage für die objektive Betrachtung von Sachverhalten, jedoch stellen sie nur eine notwendige jedoch keine hinreichende Voraussetzung für eine wissensbasierte Betrachtung von Sachverhalten dar. Das erforderliche Wissen muss zuerst durch die Interpretation und Analyse der Daten erworben werden. Durch das Aufzeigen von Zusammenhängen integrierter offener Daten können die umfangreichen Potentiale offener Daten erschlossen werden, was durch die Offenlegung der Daten alleine nicht möglich ist. Für die Generierung von Information kann aus einem breiten Spektrum von Möglichkeiten der Statistik geschöpft werden; von der einfachen Beschreibung von Daten wie der Berechnung einfacher Maßzahlen, bis hin zu komplexen statistischen Methoden der Inferenzstatistik oder der strukturentdeckenden Verfahren. Werden solche

Informationen mit fachlichem Wissen in Beziehung gesetzt, also interpretiert und kontextualisiert, entsteht aus diesen Informationen Wissen.

Je offener ODAAAaaS gestaltet wird und je größer der angebotene Funktionsumfang ist, desto eher lassen sich hier Potentiale für erfahrene Nutzer realisieren. Dabei spielt fachliches Wissen eine entscheidende Rolle, da nur so Daten sinnvoll interpretiert werden können. Um die größtmöglichen Potentiale von ODAAAaaS zu realisieren ist daher eine zumindest teilweise Kodifizierung und Formalisierung solchen Domänenwissens von Vorteil. Diese Kodifizierung kann durch die Bereitstellung spezieller Analysewerkzeuge und durch die Anreicherung der Daten mit fachlichen Metadaten erfolgen. Die in Open Data-Plattformen hinterlegten Metadaten sind dabei nur ein erster Schritt. Sie beschreiben die Daten zwar, können aber nur in geringem Maße das für die Extraktion von Wissen notwendige Domänenwissen abdecken. Eine kontinuierliche Pflege von Metadaten und das Schaffen von Anreizen auf Seiten der öffentlichen Verwaltung, Metadaten in einer hohen Qualität zur Verfügung zu stellen, sind dazu wünschenswert.

Neben dem direkten Nutzen, Daten durch die Anwendung von Analyseverfahren in Wissen zu transformieren, ergeben sich auch indirekte positive Effekte. Das Aufzeigen der Möglichkeiten offener Daten jenseits der einfachen Darstellung in tabellarischer Form oder als Kartierung von georeferenzierten Daten ist dabei ein wesentlicher Treiber, um das Thema Open Data in Wirtschaft, Wissenschaft, Politik und Gesellschaft zu etablieren. Besondere Potentiale ergeben sich vor allem aus der Kombination offener Daten mit weiteren Daten, d.h. einer Ausweitung des Integrationsschrittes auf eigene, private Daten oder auf öffentlich verfügbare, möglicherweise Big Data. Eine Vereinfachung des Umgangs mit offenen Daten durch die Bereitstellung von Analysewerkzeugen erschließt weitere Nutzerschichten. Dies führt dazu, dass die Analyse von Daten außerhalb akademischer Kontexte und Fragestellungen die Meinungsbildung innerhalb der Zivilgesellschaft verbessern kann und weitergehende Dialoge zwischen Politik und Bürger ermöglicht werden.

Eine besondere Rolle bei der Transformation von Daten zu Wissen kommt dabei der Visualisierung von Daten zu. Im Gegensatz zur Interpretation statistischer Modelle ist eine visualisierte Aufbereitung für einen wesentlich größeren Nutzerkreis zugänglich. Avancierte Visualisierungssysteme erlauben es dabei, dem Nutzer nicht nur einzelne Datensätze oder Variablen zu verstehen, sondern auch Zusammenhänge zwischen Daten zu erkennen. Besondere Potentiale ergeben sich für interaktive Visualisierungen, in denen der Nutzer die Darstellung der Daten dynamisch verändern kann.

Die Nutzung von Analysewerkzeugen kann zusätzlich die Filterung und einfache Transformation von Daten umfassen, so dass Nutzer die Daten auf Ihre Bedürfnisse hin anpassen können. Bezieht sich z.B. das Erkenntnisinteresse bei geobasierten Daten auf eine höhere Abstraktionsebene, z.B. der Zusammenfassung von georeferenzierten Daten auf Stadt oder Landesebene, so kann dies durch den Nutzer nicht auf eine einfache Art und Weise ohne den Einsatz entsprechender Werkzeuge erreicht werden.

Automatisierte Maschinenverarbeitbarkeit

Die Bereitstellung offener Daten erfolgt bereits heute zum großen Teil in maschinenlesbarer Form. Die automatische Maschinenverarbeitbarkeit offener Daten ist jedoch nicht immer gewährleistet. Dies liegt u.a. darin begründet, dass Daten zum Teil so veröffentlicht werden, dass die direkte Lesbarkeit für Menschen im Vordergrund steht. Dies ist für die Nutzbarkeit offener Daten in einem breiteren Kontext nicht zielführend und entspricht auch nicht gängigen Open Data Definitionen. Ein wesentlicher Grund für die Veröffentlichung offener Daten in für Menschen direkt verständlicher Weise liegt darin, dass nur wenige Open Data-Plattformen Funktionen zur Filterung, tabellarischen und graphischen Visualisierung bereitstellen. Eine Trennung von Daten und Visualisierungsfunktionen bietet den entscheidenden Vorteil, dass durch die Auswahl geeigneter Funktionen eine Anpassung an die spezifischen Bedürfnisse des Nutzers möglich ist. Dieses Vorgehen setzt jedoch voraus, dass die Daten mit ihren zugehörigen Metadaten direkt in maschinenverarbeitbarer Form vorliegen.

5.2 Kompetenzen der ODAaaS-Akteure

Die Untersuchung möglicher ODAaaS-Geschäftsmodelle muss aus den spezifischen Perspektiven der identifizierten Akteure erfolgen. Voraussetzung für jedes Geschäftsmodell ist, dass die offenen Daten in bearbeitbarer Form vorliegen oder entsprechend aufbereitet werden können. Ergänzend müssen bei den beteiligten Akteuren spezifische Kompetenzen vorhanden sein. Das Geschäftsmodell hängt somit vom Grad und der Ausprägung vorhandener technischer, statistischer und fachlicher Kompetenzen sowie den bei der Nutzung von ODAaaS vorliegenden technischen, organisatorischen und wirtschaftlichen Anforderungen ab. Je nach vorhandener Ausprägung von Kompetenzen und Anforderungen ergeben sich unterschiedliche Formen der Wertschöpfung.

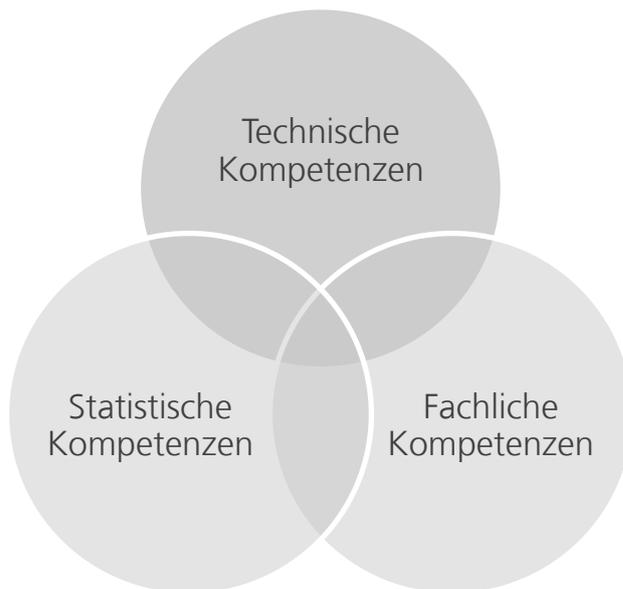


Abbildung 5: Kompetenzen der Akteure in ODAaaS

So können beispielsweise fachliche Kompetenzen bez. einzelner oder mehrerer Domänen erforderlich sein oder statistische Kompetenzen bez. einfacher, beschreibender Statistik oder bez. umfangreicher Methoden wie inferenzstatistischer, strukturentdeckender oder semantischer Simulationsverfahren erforderlich sein. Auf diese Weise können auch Geschäftsmodelle für spezialisierte Unternehmen betrachtet werden, die für spezifische Fragestellungen wettbewerbsfähige Geschäftsmodelle umsetzen können. Im Folgenden werden die identifizierten Kompetenztypen kurz beschrieben.

Technische Kompetenzen

Technische Kompetenzen beziehen sich auf die in Kapitel 4 identifizierten technischen Anforderungen sowie die in Kapitel 6 dargestellten Varianten der ODAaaS Bereitstellung als Cloud-Service. Inhaltlich stehen vor allem

- der Umgang mit Datenformaten und Metadaten aus dem Bereich Open Data,
- die Entwicklung von Werkzeugen zu Datenbereinigung, Integration und Analyse sowie
- die Bereitstellung zugehöriger Werkzeuge als Cloud-Dienste

im Vordergrund. Für den Umgang mit heterogenen Datenformaten und Metadaten (Retrieve-Schritt), Datenbereinigung (Revise-Schritt), Datenintegration (Integrate-Schritt) und Datenauswertung (Analyze-Schritt) werden zugehörige Werkzeuge benötigt. Diese Dienste und Werkzeuge werden vom Akteur „Dienstentwickler“ bereitgestellt, der die dazu benötigten Kompetenzen besitzen muss. Sollen Endnutzer ihrerseits neue Werkzeuge erstellen können, müssen sie ebenfalls über technische Kompetenzen verfügen. Die Bereitstellung der Werkzeuge als Cloud-Dienste erfordert Erfahrungen im Umgang (IaaS, PaaS) und ggf. im Betrieb von Cloud-Diensten.

Statistische Kompetenzen

Statistische Kompetenzen umfassen alle Aspekte, die mit der Durchführung statistischer Analysen zusammenhängen. Statistische Kompetenzen werden in jedem der vier Prozessschritte benötigt. Sie umfassen die Erhebung von Daten, deren Bereinigung und Integration und die Durchführung von Analysen einschließlich der Entwicklung und Interpretation von Modellen vor dem Hintergrund der statistischen Aussagekraft.

Fachliche Kompetenz

Fachliche Kompetenz beschreibt das Wissen, das zur fachlichen Interpretation von Daten, also deren Informationsgehalt und der Kontextualisierung von Analyseergebnissen relevant ist. Fachliche Kompetenz umfasst Kontextwissen, das sich auf Entstehung der Daten und relevante Fragestellungen bezieht. Sie ist im Kontext von ODAaaS nicht nur zwingend auf Seiten der Datenbereitsteller und Analysten erforderlich. Auch Entwickler domänenspezifischer, fachlicher Speziallösungen müssen über entsprechende Kompetenzen verfügen.

Grundlage der Wertschöpfungsquellen - Eigenschaften offener Daten

Neben verschiedenen Ausprägungen der betrachteten Kompetenzen ergeben sich aus den Rahmenbedingungen, welche momentan im Bereich Open Data vorliegen, entsprechende Wertschöpfungspotentiale. Dabei kann davon ausgegangen werden, dass offene Daten Informationen beinhalten, welche von Interesse für bestimmte Zielgruppen sind. Folgende Eigenschaften offener Daten können als wichtig in Bezug auf ihre Nutzung innerhalb einer Wertschöpfungskette angesehen werden.

- Offene Daten liegen bisher nur zu einem geringen Anteil integriert vor
- Offene Daten können von Qualitätsproblemen betroffen sein
- Metadaten zu offenen Daten können von Qualitätsproblemen betroffen sein
- Offene Daten ermöglichen insbesondere durch die Zusammenführung mit privaten Daten die Beantwortung spezifischer Fragestellungen

Die genannten Eigenschaften beeinflussen Geschäftsmodelle, in denen Akteuren der Kategorie „Dienst-/Datenutzer“ Daten in verbesserter Form zur Verfügung stellen. Bürger, Verwaltung und häufig auch kleinere Unternehmen verfügen nur zu einem geringen Teil über Kompetenzen bei der Datenverbesserung, sind jedoch an der Nutzung bereits veredelter Daten interessiert. Während offene Rohdaten eine kostenlose Ressource darstellen kann die Aufbereitung zu veredelten Daten je nach Lizenzmodell bereits mit einem Geschäftsmodell verbunden sein. In vielen Anwendungsfällen bilden offene Daten nur eine Teilmenge der Daten, die für Analysen genutzt werden. Die Zusammenführung offener Daten und privater Daten ermöglicht neue Analysepotentiale. Dabei kann davon ausgegangen werden, dass solche Modelle sich eher auf spezifische Fragestellungen und Datensätze beziehen. Entsprechend intelligent ausgestaltete Analysen sind dabei die Grundlage der Monetarisierung. Ergänzend muss beachtet werden, dass die bei den Datennutzern vorhandenen Kompetenzen einen wichtigen Einfluss auf die Wertschöpfungskette haben. Zudem können die Daten aus technischen, statistischen und erkenntnisorientierten Perspektiven ausgewertet werden. Diese Rahmenbedingungen beeinflussen die möglichen Betätigungsfelder, Geschäftsmodelle und konkurrierenden Dienstangebote der beteiligten Akteure.

Grundlage der Wertschöpfungsquellen - Kombination der Kompetenzen

Aus den drei beschriebenen Kompetenzklassen lassen sich konkrete Möglichkeiten ableiten, je nachdem in welcher Form die Kompetenzen vorgehalten bzw. nutzbar gemacht werden können. Im Folgenden sollen diese Möglichkeiten für die einzelnen Typen der betrachteten Akteure beschrieben werden. Zunächst werden Akteure betrachtet, welche nur über Kompetenzen aus einer Klasse verfügen. Ihre Rolle ist allgemein darauf beschränkt, diese Kompetenzen in generischer Form, z.B. über die Entwicklung von Werkzeugen, der Entwicklung statistischer Methoden oder der Entwicklung fachlicher Modelle in ODAaaS einzubringen. Ihr Beitrag ist dennoch aus Sicht

einer Gesamtkonzeption nicht zu unterschätzen. Sie können in vielfältiger Weise der Unterstützung anderer Akteure dienen, welche stärker in der Entwicklung von ODAaaS involviert sind.

Für Akteure mit zwei Kompetenzen eröffnen sich bereits aktive Gestaltungsspielräume. Je nachdem in welcher Kombination die Kompetenzen vorliegen sind sie eher Nutzer oder Anbieter von ODAaaS-Diensten.

Liegen Kompetenzen in allen drei Klassen vor, so kann von sehr hohem Potential für die wirtschaftliche Verwertung ausgegangen werden. Hierbei ist die Unterscheidung in umfassende oder spezialisierte Ausprägung der Kompetenzen entscheidend für die Wahl sinnvoller Betätigungsfelder.

Im Folgenden werden die charakteristischen Eigenschaften der Reinformen von Kompetenzträgern beschrieben. Dabei wird das Ziel einer Zuordnung von Akteuren zu Betätigungsfeldern und Geschäftsmodellen verfolgt. Der Umfang der vorhandenen Kompetenz ist dabei zunächst unerheblich und empirisch auch kaum umfassend zu erschließen. Auch ist es durchaus plausibel, dass Akteure ihre Kompetenzen erweitern können oder mit Akteuren, welche über fehlende Kompetenzen verfügen, kooperieren. Das Ziel dieser Charakterisierung ist das Erkennen von Potentialen einer möglichst breiten Gruppe von Akteuren in Abhängigkeit von deren Kompetenzen. Die unterschiedlichen Kompetenzträger werden im Folgenden zunächst kurz beschrieben. Anschließend werden typische Kombinationen von Akteuren und Kompetenzen betrachtet.

Im Sinne der betrachteten Fragestellung von ODAaaS sind insbesondere diejenigen Akteure von Interesse, die domänenspezifische Wissen zur Analyse offener Daten einbringen wollen bzw. die Cloud-basierte Dienste zur Unterstützung dieser Aufgabe anbieten können.

Statistiker

Der Typus des reinen Statistikers besetzt nur eines der skizzierten Kompetenzfelder, statistische Kompetenzen, und verfügt über keine weiterführenden technischen Kompetenzen und über kein im Kontext offener Daten relevantes fachliches Wissen. Er ist für ODAaaS hauptsächlich Impulsgeber. Seine Rolle ist für die aktive Ausgestaltung von ODAaaS dahingehend relevant, dass sie einen Zufluss statistischer Methoden sicherstellt, Statistiker haben ein umfangreiches Wissen, was mit statistischen Methoden prinzipiell unabhängig von einer konkreten Fragestellung realisiert werden kann. Ihr fehlendes fachliches Wissen schränkt ihre Möglichkeiten ein, fachliche Schlüsse aus statistischen Analysen zu ziehen. An Märkten treten sie als Anbieter auf, indem sie neue Methoden entwickeln. Allgemein sind Statistiker auch Kunden von IaaS oder PaaS-Angeboten, welche es Ihnen erlauben, ihre Algorithmen in einer bedarfsgerechten IKT-Infrastruktur anzubieten.

Dieser Typus entspricht am ehesten Forschern und Forschungsinstituten, welche neue statistische Verfahren oder Visualisierungsformen entwickeln. Akteure eines solchen Typus sind primär an der Entwicklung von Methoden interessiert.

Entwickler und Techniker

Entwickler und Techniker stellen den zweiten Typus von Akteuren dar, welche nur über eine der drei Kompetenzen verfügen. Sie sind im Rahmen von ODAaaS sowohl Impulsgeber als auch aktiver Bereitsteller generischer Plattformen, Werkzeuge oder Dienstleistungen aus dem Bereich der IKT/Cloud-Infrastruktur. Ihre Rolle ist im Kontext von ODAaaS die eines technischen Anwendungsentwicklers bzw. eines Anbieters von Cloud Diensten. Sie können generische Lösungen und Werkzeuge zur Behandlung von Daten bereitstellen.

Fachliche und Domänenexperten

Reine Domänenexperten verfügen über umfangreiches Wissen innerhalb ihrer spezifischen Domäne. Sie verfügen über spezifisches Wissen, bezüglich Fragestellungen und Wissen zur Kontextualisierung von Analyseergebnissen. Sie nutzen bei ihren Analysen fertige Werkzeuge mit einem hohen Grad an Nutzerführung und Visualisierungen. Als Anbieter können sie über die Monetarisierung ihres Domänenwissen auftreten, d.h. insbesondere fachlich

aufbereitete Daten zur Verfügung stellen oder ihr aus Analysen gewonnenes Wissens Dritten anbieten. Als Nachfrager treten sie sowohl als Nachfrager für Werkzeuge aber auch als Nachfrager statistischer Analysen oder Lösungen auf.

In Teilen entspricht dieser Typ, zumindest für den Fall offener Verwaltungsdaten, Akteuren aus der öffentlichen Verwaltung aber auch interessierten und gut informierten Anwendern. Akteure diesen Typus sehen Daten als Grundlage für die Generierung von Wissen.

Statistische Programmierer

Statistische Programmierer haben sowohl ein Wissen um statistische Methoden als auch um deren Umsetzung in Werkzeuge und Dienste. Sie haben die Fähigkeit neue Verfahren zu entwickeln, auch wenn diese Fähigkeit gegenüber reinen Statistikern geringer ausgeprägt sein kann. Sie sind aber in der Lage, bestehende Algorithmen und Verfahren in einer sinnvollen Weise technisch zugänglich zu machen. Der Mangel an Domänenwissen ermöglicht ihnen aber keine direkte Wertschöpfung im Sinne der Durchführung statistischer Analysen. Sie sehen in offenen Daten Anwendungspotentiale aus ihrem Alltagsverständnis. Ihr Zugang zu offenen Daten ist daher größtenteils technischer Natur.

Technische Dienstleister mit fachlicher Kompetenz

Derartige Dienstleister verfügen über technisches Wissen und verfügen über fachliche Kompetenz in einer oder mehreren Domänen. Sie betreiben üblicherweise dedizierte Infrastrukturen und bieten darin Fachdienste an. Aufgrund fehlender statistischer Kompetenz sind sie nur eingeschränkt in der Lage, Analysen auf Basis offener Daten durchzuführen. Ihr Zugang zu offenen Daten ist domänenorientiert und eher technischer Natur. Sie sind jedoch aufgrund der Kombination aus technischem Wissen und Domänenwissen in der Lage, innerhalb einer Domäne Daten in einer sinnvollen Weise zu integrieren, so dass diese statistischen Analysen zugänglich gemacht werden können. Sie können im Kontext ODaaS entsprechend als Zulieferer integrierter Daten betrachtet werden. Als Nutzer sind sie insbesondere an statistischen Analysediensten interessiert, um zusätzlichen Nutzen aus den betrachteten Daten zu ziehen. Dieser Typus entspricht am ehesten den Anbietern von hochspezialisierten IT-Dienstleistungen innerhalb einer Domäne.

Domänenspezifische Analysten

Domänenspezifische Analysten verfügen über fachliche Kompetenz und Wissen im Bereich der statistischen Analyse. Sie haben keine technische Kompetenz, um eigene Werkzeuge zu entwickeln, sind aber in der Lage, auf Basis von Daten domänenspezifisches Wissen für Kunden zu generieren. Ihr Zugang zu offenen Daten ist sehr erkenntnisorientiert. Daten, welche für die statistische Analyse verwendbar sind werden durch sie nutzenstiftend im Rahmen komplexer Analysen verwendet. Ihre Nachfrage gilt vor allem generischen Analysewerkzeugen Werkzeugen und integrierten Daten, um ihre Stärken einzubringen. Aufgrund ihres Domänenwissens kommen sie auch als Nutzer von Werkzeugen aus dem Bereich der Datenbereinigung und Integration in Betracht. Akteure dieses Typus sind vor allem Beratungsunternehmen, aber auch Datenjournalisten oder fachlich orientierte Forschungsinstitute.

Domänenspezifische Anbieter von Analysediensten

Domänenspezifische Lösungsanbieter und Integratoren verfügen innerhalb ihrer Domäne über technisches und statistisches Know-how und können ihr Wissen so kombinieren, dass sie für interessierte Kreise spezifische Kundenlösungen realisieren können. Ihre Sicht auf Daten ist tiefgehend. Sie erkennen das analytische Potential der Daten und bieten Datenanalysen und generiertes Wissen in Form von Dienstleistungen an. Je nach vorhandener Ressourcenausstattung können diese Akteure weiterhin zugehörige Werkzeuge und Dienste entwickeln und diese, gegebenenfalls als mandantenfähige SaaS-Dienste, zur Verfügung stellen.

Derartige Akteure können Basistechnologien oder neue Verfahren im Sinne einer Open Innovation Strategie integrieren. Sie legen einen starken Fokus auf den Zufluss von Wissen und Technologien. Am Markt treten sie sehr stark am Ende der Wertschöpfungskette auf und kombinieren bestehende Technologien.

Laiennutzer

Bei Laiennutzer liegt kein oder nur sehr wenig Wissen und Kompetenzen aus den einzelnen Bereichen vor. Sie verfügen nur eingeschränkt über Domänen- oder Statistikwissen, um komplexe Analysen durchzuführen oder deren Ergebnisse zu interpretieren. Hingegen können sie einfache Auswertungen und Visualisierungen offener Daten durchführen und ggf. als Teil einer Entwickler-Community über zugehörige Apps verfügbar machen. Sie treten am Markt zumeist als Nutzer auf, können aber in Crowdsourcing-Ansätzen durchaus bei der Veredlung offener Daten mitwirken oder eigene Daten bereitstellen.

Sie fragen selbst nur einfache Werkzeuge nach und sind primär an Analyseergebnissen Dritter interessiert. Ihr Zugang zu domänenspezifischen Wissen ist an Fragestellungen orientiert, die sich aus ihrem Alltagsverständnis oder konkreten Anforderungen ableiten. Der Zugang zu Daten und Statistiken ist intuitiver Natur und sie bedürfen Unterstützung bei der Interpretation komplexer Analysen. Die kritische Interpretation von Indikatoren oder Maßzahlen stellt sie im Einzelfall vor hohe Anforderungen. Am ehesten lassen sich solche Nutzer durch die Visualisierung von Daten erreichen. Da ihnen der Zugang zur Komplexität analytischer Verfahren und dem Aufwand an Rechenzeit fehlt, zeichnen sich solche Nutzer dadurch aus, dass sie Information sehr schnell innerhalb kürzester Zeit erhalten wollen und nur wenig Einfluss auf den eigentlichen Analyseprozess nehmen wollen. Einfachheit in der Präsentation und Nutzerfreundlichkeit steht für sie im Vordergrund.

5.3 Umsetzungskonzepte für ODAaaS

Anhand der in Kapitel 4 ausgeführten unterschiedlichen Ausgestaltungsformen von ODAaaS und der bei einzelnen Akteuren unterschiedlichen Kombination von Wissen und Kompetenzen lassen sich den Akteuren spezifische Bestandteile und Prozessschritte von ODAaaS zuordnen. Diese Zuordnung schließt jedoch nicht aus, dass Akteure jederzeit weitere Kompetenzen aufbauen und sich neues Wissen aneignen können. Auch werden einzelne Prozessschritte nicht nur von einer Akteursgruppe besetzt. So können beispielsweise domänenspezifische Anbieter von Analysewerkzeugen alle Prozessschritte durchführen. In welcher Breite sich solche Lösungsanbieter positionieren obliegt dem einzelnen Akteur. Im Folgenden werden die in unterschiedlichen Prozessschritten benötigten bzw. angebotenen Dienste geeigneten Akteuren zugeordnet.

Fachspezifische Information as a Service-Dienstangebote

Fachspezifische Information as a Service-Dienste stellen Daten in integrierter Form für Nachfrager in einer für deren Zwecke geeigneten Form über webbasierte Schnittstellen zur Verfügung. Hierbei sind Information as a Service-Anbieter auf eine Beschreibung der Daten auf Basis der Metadaten der Datenvektoren angewiesen oder müssen diese im Zuge der Bereitstellung selbst erstellen. Im letzten Fall besteht ein erhöhter Standardisierungsbedarf in Bezug auf die verwendeten Metadatenstandards.

Ein einfaches Anwendungsbeispiel besteht in der Zusammenfassung integrierter Daten nach geographischen Kriterien. Komplexere Beispiele liegen vor, wenn die semantische Beschreibung keine eindeutige Zuordnung und Zusammenführung zulässt.

Bei der Integration der Daten werden entweder Ontologien zur automatisierten Zusammenführung oder ein tiefgehendes Domänenwissen zur manuellen Zusammenführung benötigt. Mangelnde Qualität der zusammengeführten Ausgangsdaten schlägt sich auf die integrierten Daten durch. Der Nutzen der Information as a Service-

Dienste besteht darin, dass die Kosten der Bereinigung und Integration von Nachfragern nicht mehr selbst erbracht werden müssen. Die für diese Dienstleistungen benötigten Ressourcen fallen je nach Datenbeschaffenheit unterschiedlich aus. Bei der Behandlung sehr großer Datenbestände im Sinne des Big Data ist beispielsweise ein umfangreiches technisches Know-how aus dem Bereich der Analyseverfahren eine wesentliche Grundvoraussetzung.

In einer Cloud-Umgebung können Information as a Service-Anbieter einerseits als Anbieter von Daten auftreten, die sie über zugehörige Dienstschnittstellen verfügbar machen. Speichern sie die integrierten Daten selber innerhalb einer Cloud, so treten sie dabei als IaaS-Nutzer auf. Ihr typischer Kunde ist der „Data Analyst“. Andererseits können Information as a Service-Anbieter auch als Anbieter von PaaS-Lösungen auftreten. In diesem Fall bieten sie eine Analyseplattform an, in die verschiedene Analysedienste eingebettet werden können und dort auf integrierte Daten zugreifen. Ihr typischer Kunde ist der „Analytics Service Provider“. Für Bürger in der Rolle „Data Analyst“ ermöglicht der Nutzen von „Information as a Service“ beispielsweise durch gezielte und einfache Beschäftigung mit integrierten, offenen Daten die Verbesserung von Bürgerinformationssystemen.

Open Data-Portale stellen in der Regel einen Datenkatalog zur Verfügung. Sie beinhalten zumeist keine weitergehende Funktionalität für die Durchführung weiterer Prozessschritte. Eine Erweiterung von Open Data-Portalen ist als eine Art Premiumdienst vorstellbar, welcher direkt aus einem Open Data-Portal über Verlinkung zu den entsprechenden Diensten realisiert werden kann. Kritisch ist bei solchen Diensten die Behandlung der Dienstqualität. Selbst wenn die Dienstqualität technisch beschrieben werden kann, ist sie für den Typus des Nutzers im Einzelnen nur schwer nachvollziehbar. Dies kann dazu führen, dass Nutzer verunsichert werden und nur bedingt bereit sind, für solche Dienste hohe Preise zu entrichten.

Kritisch muss auch gesehen werden, dass nachgelagerte Dienste mit den in Open Data-Portalen referenzierten Datenbeständen harmonisiert vorliegen müssen. Dies entzieht sich im Falle einer verteilten Dienstarchitektur der Kontrolle des Information as a Service-Dienstanbieters. Liegen in Open Data-Portalen veraltete Zugriffspunkte vor, sind also Daten nicht mehr verfügbar, so müssen Anbieter dies bei ihrem Dienstangebot berücksichtigen.

Ontologien zur Datenintegration

Ontologien spielen bei der Zusammenführung von Daten im Rahmen von ODAaaS eine wichtige Rolle, da durch Ontologien Querbezüge zwischen Daten hergestellt werden können. Im Kontext von ODAaaS können Ontologien auf zwei Arten zum Einsatz kommen. Erstens, können Ontologien mittelbar über die Herstellung von Querbezügen zwischen Daten als Instrumente der Datenintegration verstanden werden. Zweitens, können Ontologien als Werkzeug bei der Transformation von Daten verstanden werden. Ontologien stellen einen Input für Information as a Service dar und können als Bindeglied zwischen Information as a Service und AaaS eingesetzt werden. Die Erstellung von Ontologien erfordert dabei umfangreiches Domänenwissen. Nutzer, die zwar über solches Domänenwissen verfügen, aber keine technischen Kompetenzen besitzen, müssen dabei durch entsprechende Werkzeuge unterstützt werden.

Allgemeine AaaS-Dienstangebote

Die Kernidee hinter ODAaaS besteht in dem Ansatz, Analysedienste über offenen Daten als Cloud-Dienste anzubieten. Ein AaaS-Dienstanbieter bietet allgemein Analysedienste als Cloud-Dienst an, ohne dabei auf die Besonderheiten offener Daten einzugehen. Anbieter solcher Dienste müssen nicht über domänenspezifisches Wissen verfügen. Ihre Aufgabe im Rahmen von ODAaaS beschränkt sich auf die Bereitstellung des Dienstes selbst. Sie tragen dafür Sorge, dass statistische Verfahren in mathematisch korrekter Weise implementiert sind und dass ein Zugriff auf über gängige Methoden referenzierte offene Daten möglich ist.

AaaS-Dienstleistungen setzen eine angemessene Qualität der zu analysierenden Daten voraus oder benötigen selber vorgelagerte Importfunktionen zur Verbesserung der Datenqualität und zur Integration von Daten aus

verschiedenen Quellen. Allgemeine AaaS-Systeme müssen daher entweder entsprechende externe Dienstleistungen in Anspruch nehmen oder aber selber als Teilfunktionalität zur Verfügung stellen. Im Sinne wiederverwendbarer Cloud-Dienste bietet sich für ODAaaS die erste Lösung mit speziellen Importdiensten für aufbereitete offene Daten an. Anbieter von AaaS-Diensten müssen in der Lage sein, statistische Operationen und Visualisierungen in Form eines webbasierten Dienstes zur Verfügung zu stellen. Sie müssen dabei gleichzeitig über hinreichendes statistisches Wissen verfügen, um die statistischen Methoden in richtiger aber auch sinnvoller Weise zur Verfügung zu stellen. Sie benötigen nur wenig Domänenwissen, insofern sie Daten nicht prozessieren, sondern lediglich die Darstellungs- und Analyseebene offener Daten bedienen.

Die Beschreibung einer AaaS-Dienstleistung besagt, welche Operationen auf Daten angewendet werden, welche Verfahren auf Daten angewendet werden, welche Daten für solche Operationen zulässig sind und in welcher Weise das Ergebnis zur Verfügung gestellt werden kann. Die größte Anforderung besteht dabei in einer umfassenden Beschreibung der Charakteristika zulässiger Daten bzw. der unterstützten Importdienste sowie einer allgemein verständlichen Darstellung der Ergebnisse und deren Interpretierbarkeit.

Indikatorentwicklung

Indikatoren dienen der Messung latenter und nicht direkt beobachtbarer Phänomene. Ein Beispiel dafür ist die Messung von Innovationsoutput über Indikatoren auf Basis von Patentanmeldungen. Indikatoren sind nicht in allen Domänen gleich aussagekräftig. Die Entwicklung von Indikatoren setzt ein umfangreiches Wissen über die verwendeten Daten aber auch über deren Kontextualisierung im Rahmen von Domänen dar. Sie sind meist das Produkt einfacher statistischer Verfahren und können auf mehreren Datensätzen aufbauen.

Im Rahmen von ODAaaS sind domänenspezifische Analysten in der Lage, Indikatoren zu konstruieren, da sie sowohl über statistisches Wissen aber auch entsprechendes Domänenwissen verfügen. Unterstützt eine allgemeine AaaS-Lösung die Definition fachspezifischer Indikatoren, kann sie relativ einfach zur Beantwortung fachspezifischer Fragestellungen herangezogen werden.

Fachspezifische AaaS-Dienstangebote

Fachspezifische AaaS-Dienstangebote zeichnen sich gegenüber allgemeinen AaaS-Dienstangeboten durch eine fachliche Spezialisierung der Analysefunktionen für spezielle Anwendungsdomänen aus. Im Sinne der gegebenen Problemstellung existieren keine weitergehenden Anforderungen gegenüber den allgemeinen Lösungen. Man kann davon ausgehen, dass der Nutzerkreis eingeschränkt wird, da die Ergebnisse nur für fachliche Spezialisten verständlich sind. Andererseits sind die Ergebnisse dafür aber aussagekräftiger und für Spezialisten leichter ermittelbar und interpretierbar, so dass die Anzahl der Fachnutzer wiederum gegenüber allgemeinen Lösungen steigen dürfte.

Visualisierung von Analyseergebnissen durch mobile Apps

Eine weitere wirtschaftlich wie technisch relevante Anforderung besteht in der Nutzbarkeit von ODAaaS auf mobilen Endgeräten. Der Ansatz, ODAaaS als Cloud-Dienst anzubieten besagt per Definition nur, dass ODAaaS-Funktionen über Web-Schnittstellen in Anspruch genommen werden können. Ob dies über schwergewichtige Klienten, GUIs in Web-Browsern oder über mobile Apps geschieht ist eine Frage der technischen Umsetzung. Je nach spezifischen Anforderungen und dem Grad der notwendigen Interaktivität können mobile Endgeräte, insbesondere Smartphones, aufgrund der Größe der zur Verfügung stehenden Displays nur bedingt herangezogen werden. Für auf klar umrissene Fragestellungen zugeschnittene, fachspezifische Analyse-Werkzeuge (Verkehrsdichte, Fahrpläne, Luftverschmutzung, (Un-)Wetterprognosen usw.) stellen jedoch Smartphone-Apps eine sinnvolle und populäre Lösungsvariante dar. Die grafische Darstellung auch komplexere Endergebnisse einer Datenanalyse auf Tablets ist in vielen Fällen eine sinnvolle Alternative. Der Endnutzer ist bei diesen Beispielen nicht an der Durchführung der Datenanalyse sondern ausschließlich an der Visualisierung der Ergebnisse interessiert.

6 Expertenforderungen an ODAaaS

6.1 Methodische Vorbemerkungen

Im Folgenden soll die Frage diskutiert werden, in welcher Form ODAaaS, d.h. die cloudbasierte Analyse offener Daten, umgesetzt und genutzt werden kann. Hierzu ist zunächst zu prüfen, in welcher Form die in Kapitel 4 abgeleiteten Anforderungen und grundlegenden Gestaltungsspielräume auf Basis der heute am Markt befindlichen Lösungen anwendbar sind und von Experten als sinnvoll angesehen werden.

Um mögliche Potenziale zur Etablierung von ODAaaS sowie Anforderungen von technisch möglichen Ausgestaltungsformen zu beurteilen, müssen die heutige Nutzung des Funktionsumfangs von Open Data Plattformen und die Anforderungen auf Seiten der Datenbereitstellung betrachtet werden. Im Rahmen dieser Studie wurden Experteninterviews mit Vertretern von fünf Unternehmen aus dem Bereich Cloud und AaaS sowie mit zwei Experten nicht kommerzieller AaaS-fähiger Open-Source-Lösungen durchgeführt. Dabei sollte untersucht werden, welche Anforderungen und Potentiale sich insgesamt für die Integration von Open Data und AaaS ergeben. Im Vordergrund steht eine synthetische Betrachtung der Expertenaussagen. Die Interviews wurden teilstrukturiert durchgeführt.

6.2 Ergebnisse der Experteninterviews

Für die Mehrzahl von Lösungen der Interviewpartner stellen die technischen Anforderungen kein Problem dar. Dabei wurden einige Anforderungen jedoch nur in Teilen berücksichtigt. Die Aussagen der Interviewpartner lassen darauf schließen, dass grundlegende Lösungsansätze für mögliche Ausgestaltungsformen von ODAaaS vorhanden sind. Meist werden Lösungen kundenspezifisch aufgebaut und mit entsprechenden Beratungsdienstleistungen hinterlegt.

Allgemein verfügen alle Interviewpartner über entsprechende Lösungen, um Datenanalyse in der Cloud als SaaS auszuführen. Hierbei werden Datenanalysedienste meist vollständig aus dem Leistungsportfolio der Unternehmen selbst abgedeckt. Im Einzelfall werden Teilaspekte über die Integration von Partnerlösungen etabliert. Dies ist zum Beispiel im Bereich der Abrechnung von Diensten der Fall.

In Bezug auf eine stärker verteilte Lösung stehen derzeit technische Lösungen zur Datenintegration als Information as a Service-Dienste und für Speicherung von und Zugriff auf Daten als Storage as a Service-Dienste zur Verfügung. Standardisierte Schnittstellen und Konzepte für die Kommunikation von Diensten untereinander und deren Organisation liegen ebenfalls bei einigen Anbietern vor und stellen z.T. einen wesentlichen Anteil ihrer Geschäftsmodelle dar. Sie werden jedoch überwiegend im Kontext spezifischer Kundenlösungen angeboten.

Die üblicherweise im Zuge der Veröffentlichung offener Daten verwendeten Formate und Schnittstellen werden sowohl von kommerziellen als auch nicht-kommerziellen Lösungen unterstützt. Alle kommerziellen Datenanalyselösungen verfügen über umfangreiche Funktionalitäten zur Integration und Nutzung heterogener und verteilter Daten in Open Data-Plattformen. Ihre Nutzung kann laut Aussage von Experten jedoch mit einem erheblichen Aufwand verbunden sein, insbesondere dann, wenn die Daten in unzureichender Qualität vorliegen und kein Domänenwissen zur Verfügung steht oder durch den Kunden nicht in adäquater Form eingebracht werden kann.

Die befragten Unternehmen bauen in der konkreten Umsetzung, je nach spezifischen Kundenanforderungen, entweder auf dedizierte eigene Lösungen oder auf entsprechende Partnerlösungen auf.

6.2.1 Metadaten

In Bezug auf Metadaten werden durch nahezu alle Lösungen entsprechende Katalogstandards unterstützt oder können nach Angaben zeitnah implementiert werden, so dass ein automatisiertes Abfragen der in der Open Data Plattform hinterlegten Daten technisch möglich sein sollte. Bezüglich der Datenprovenienz werden unterschiedliche Ansätze gewählt. Dabei stehen je nach Lösung interne oder externe Strategien im Vordergrund. Einige Systeme beziehen die Provenienz eher auf die Nutzer ihres Systems und verstehen diese Nutzer als Besitzer oder Bereitsteller der Daten. Die Provenienz der Daten wird dabei über die Rolle des Datenbereitstellers als „Datenquelle“ gesehen. Andere Konzepte trennen Datenquelle und Datenbereitsteller nicht in dieser Weise auf, sondern integrieren sie konzeptionell in einem den Daten zugeordneten Metadatensatz.

In der Praxis spielt dies jedoch eine nachgelagerte Rolle, insofern die eigentliche Information über die Datenquelle in beiden Fällen erhalten bleibt. Bei der weiteren Bereitstellung der Daten kann diese Unterscheidung abstrahiert werden, sofern einheitliche Metadatenstandards gewählt werden, bzw. die entsprechenden Informationen in Metadaten abbildbar sind.

Allgemein bieten alle Lösungen die Möglichkeit, die Verbindung von Daten und Metadaten über alle Prozessschritte zu halten, so dass in jedem Prozessschritt sowohl auf Daten als auch auf Metadaten zugegriffen werden kann. Bei den kommerziellen Lösungen sind weiterhin interne Prozesse implementiert, welche Protokolle über die Veränderung der Daten durch Transformationen bereitstellen. Je nach Lösung stehen jedoch unterschiedliche Ansätze im Vordergrund. Bei einigen modularen Systemen aus vorgefertigten und nicht durch den Nutzer veränderbaren Prozessen wird lediglich protokolliert, dass ein entsprechendes Datum oder ein Datensatz durch ein Modul verarbeitet wurde. Weiterführende Konzepte erlauben eine Protokollierung, welche Prozeduren die Daten verarbeitet haben.

6.2.2 Barrierefreiheit

Barrierefreiheit wird in marktgängigen Lösungen in unterschiedlicher berücksichtigt. Generell werden bei Komponenten, die dem Nutzer direkt zur Verfügung stehen, Mindeststandards umgesetzt, wie z.B. Teile der ISO-9241-Normenreihe. In einigen Fällen werden auch Module für die sprachgesteuerte Bedienung angeboten. Je nach Ausgestaltung der AaaS-Lösung würde sich eine sprachgesteuerte Nutzerführung jedoch als schwierig erweisen. Dies gilt insbesondere, wenn die Anwendung mit einem hohen Maß an Interaktivität verbunden ist, z.B. dem Filtern relevanter Daten aus größeren Datenbeständen. Interoperabilität stellt für die Umsetzung kein Hindernis dar, insofern sie auf der technischen Ebene zwischen der SaaS-Anwendung und dem einzusetzenden Web-Browser gewährleistet ist.

6.2.3 Verfügbarkeit und Aktualität offener Daten

Aktualität und Verfügbarkeit von Daten werden von allen hierzu befragten Interviewpartnern als wesentlich für den Erfolg eines ODAaaS-Ökosystems angesehen. In Fällen, in denen Daten weitgehend automatisiert in vorgegebenen Zeitabständen abgefragt werden, ist die Sicherstellung der Datenverfügbarkeit ein wesentlicher Faktor. Hier sehen die Experten unterschiedliche Verantwortlichkeiten um die Datenverfügbarkeit zu gewährleisten. Einige sehen die Datenbereitsteller als zentrale Akteure in der Verantwortung. Ihre Aufgabe sollte

es sein, Open Data-Portale bei Änderungen von Datensatzzugriffspunkten entsprechend zu aktualisieren. Es wird jedoch auch argumentiert, dass dies ein erhebliches Koordinationsproblem darstellen kann. Eine mögliche Lösung sehen einige Experten in der Nutzung von zentralen Speicherkonzepten oder Data Warehouses. Hierdurch reduziere sich die Problematik, dass Daten möglicherweise nicht mehr verfügbar sind oder URLs sich möglicherweise verändert haben. Der Nachteil solcher Lösungen liege jedoch darin, dass sie sich negativ auf die Aktualität der referenzierten Daten auswirken könnten. So kann es laut einem Interviewpartner zu einem Zielkonflikt zwischen Aktualität und Pflegeaufwand kommen. Für den Betrieb solcher Plattformen müsse insbesondere die Datenpflege genau, z.B. durch Verträge mit Dienstleistern, geregelt werden. Alternativ müssten klare organisatorische Regeln zu diesem Zweck implementiert werden. Dies mache die Rolle eines Intermediärs notwendig.

Was für die Datenverfügbarkeit gilt, gilt in abgeschwächter Form auch für die Aktualität von Daten. Während im Falle der Sicherstellung der Verfügbarkeit vor allem verhindert werden soll, dass Nutzer den Dienst als „unausgegoren“ empfinden, rückt bezüglich der Aktualität von Daten eher der Nutzen als solcher in den Vordergrund. Analysen auf Basis veralteter Daten stellen aus Sicht einiger Interviewpartner – je nach Verwendungskontext – das Risiko für den Betreiber dar, dass die Glaubwürdigkeit der aus der Analyse gewonnenen Erkenntnisse reduziert wird.

Allgemein liege für die Aspekte Aktualität und Verfügbarkeit insofern ein Kollektivgutproblem vor, dass Unklarheit darüber besteht, welche Akteure zur Sicherstellung dieser beiden Aspekte herangezogen werden können. Allgemein wird dies entweder im Aufgabenbereich der Datenbereitstellung oder im Bereich der Betreiber von Open Data-Plattformen gesehen. Gleiches gilt auch für den bereits besprochenen Umfang des Datenkatalogs von Open Data-Portalen und der Qualität der Daten.

6.2.4 Datenqualität

Nach Ansicht der Experten steht die Datenqualität für alle Ausgestaltungsformen von ODAAAAS an erster Stelle und muss so früh wie möglich im gesamten Verwertungsprozess realisiert werden. Dabei wird erwähnt, dass ein Teil der Anforderung darin besteht, dass moderne Instrumente der Datenbereitstellung in der öffentlichen Verwaltung noch nicht flächendeckend eingesetzt werden. Um die Ausgangslage für ODAAAAS zu verbessern, sollten solche Instrumente in der öffentlichen Verwaltung zeitnah eingebunden und Mitarbeiter entsprechend geschult werden. Die Anwendung von SaaS wird dabei z.T. als kritisch angesehen. Daten sollten bis zur Veröffentlichung innerhalb der Organisation verbleiben und auch nur dort behandelt werden. Ein Einsatz von SaaS-Lösungen in diesem Bereich sei nur auf besonders sicheren Infrastrukturen, z.B. in BSI-zertifizierten Cloud-Architekturen, zu empfehlen. In diesem Kontext werden auch Medienbrüche als eine weiterhin bestehende Anforderung gesehen, da sich durch diese die Fehleranfälligkeit und der Aufwand der Datenbereitstellung erhöhen können.

In Bezug auf die Datenqualität und insbesondere bei der Abhängigkeit von der Verfügbarkeit der Daten sind auf Seiten der Hersteller umfangreiche Lösungen im Angebot. Probleme bestehen laut einem Interviewpartner auf organisatorischer und prozessspezifischer Ebene. Gerade im Fall von Veränderungen der Datenstruktur im Zuge einer Aktualisierung oder bei der Bereitstellung maschinell lesbarer aber nicht automatisiert bearbeitbarer Daten können Probleme entstehen, die durch rein technische Lösungen nicht vollständig gelöst werden können.

Im Weiteren spielt die zeitlich veränderliche Zuordnung von Daten eine wichtige Rolle. Prinzipiell ist der dynamische Charakter in Katalogstandards wie DCAT bereits angelegt und stellt hiermit per se kein technisches Problem dar. Laut einem Interviewpartner müssen bei sich verändernden Datensätzen einfach gesprochen grundlegend zwei Dinge beachtet werden:

Erstens muss bei dynamischen Daten der Zyklus der Datenerhebung und insbesondere der Datenaktualisierung bekannt sein. Bezieht sich z.B. eine Datensammlung und Veröffentlichung auf ein Haushaltsjahr, so muss klar sein, in welcher Weise diese Daten veröffentlicht werden und ob im Einzelfall Altdaten archiviert und an anderer Stelle zur Verfügung gestellt werden müssen. Werden Datenquellen relativ referenziert, handelt es sich bei der Datenquelle also um einen Datensatz mit der Gültigkeit für das aktuelle Haushaltsjahr und ist die Hinterlegung dieser Daten immer über die gleiche URL adressierbar, so können Daten vergangener Jahre eventuell nicht mehr verfügbar sein oder durch neue Daten überschrieben werden.

Zweitens muss bei dynamischen Daten die Zusammenführbarkeit gewährleistet sein. Dies bedeutet, dynamische Daten können auch dann zusammengeführt werden, wenn sich entsprechende Klassifikationssysteme, wie z.B. Wirtschaftszweigklassifikationen oder auch Klassifikationen der Haushaltsstatistik, über die Zeit hinweg verändern. Dies wiederum bedeutet, dass ggf. neue Ontologien aktualisierter Klassifikationssysteme berücksichtigt werden müssen. Liegen Daten in unterschiedlichen Klassifikationssystemen vor, sei es schwierig, diese vollständig zu integrieren.

Eine weitere Möglichkeit, mit der Qualität von Daten und Metadaten umzugehen, ist die Einrichtung einer zentralen Datenclearingstelle, welche offene Daten auf Qualitätsmängel und Metadaten auf Vollständigkeit hin überprüft. Eine solche Stelle müsste zwischen Datenbereitstellung und Veröffentlichung innerhalb eines Open Data-Portals angesiedelt sein. Vorstellbar wäre hier, dass dies durch die Betreiber von Open Data-Portalen als Teilaufgabe übernommen wird. Die Anforderung wäre jedoch, dass umfangreiches Domänenwissen vorgehalten werden muss, um eine adäquate Bewertung der Daten und Metadaten zu gewährleisten. Ein Teil der Interviewpartner sehen eine solche Datenclearingstelle eher als nicht praktikabel, da entsprechende Zertifizierungskonzepte erst erarbeitet werden müssten.

6.2.5 Hohes Potential durch Big Data und Sensordaten

Insgesamt sehen alle Interviewpartner bezüglich der Potentiale offener Daten, sowohl in ihren eigenen Unternehmen als auch global eine deutlich zunehmende Relevanz. Der Vorteil von offenen Daten sei darin begründet, dass weniger Sicherheits- oder Datenhoheitsbedenken bestehen. Dies gilt nicht, wenn nicht offene Daten mit offenen Daten in Beziehung gesetzt werden. Potentiale werden vor allem für große offene Datenbestände (englisch: Big Open Data), Sensordaten und Echtzeitdaten als besonders hoch eingestuft. Hierbei wird auch erwähnt, dass der Anteil an solchen Daten in Open Data-Plattformen noch relativ gering ist. Interviewpartner, welche in internationalen Kontexten bereits Lösungen im Rahmen offener Daten angeboten haben, weisen darauf hin, dass gerade bei dynamischen Echtzeitdaten die Verwertungsaktivitäten stark ausgeprägt sind. Ein Interviewpartner argumentiert, dass sich interessante Daten entweder durch einen besonderen Umfang, eine hohe Dynamik oder durch Verschränkung mit anderen Daten auszeichnen, da Analysewerkzeuge dem Nutzer hier, verglichen mit Rohdaten, einen wesentlichen Mehrwertgewinn erlauben. Über die Frage, ob sich die Menge an als Big Data zu bezeichnenden offenen Datenbeständen in Zukunft erhöhen wird, liegt keine Einigkeit vor. Einige Akteure gehen fest davon aus, dass zumindest im Bereich des Nahverkehrs und des Katastrophenschutzes solche Daten in Zukunft vermehrt bereitgestellt werden. Diese Einschätzung wird durch aktuelle Entwicklungen untermauert. Andere weisen auf die bisher eher zurückhaltende Offenlegung großer Datenbestände hin und sind tendenziell skeptisch, was die weitere Offenlegung solcher Daten betrifft.

6.2.6 Die Rolle einer Community bei der Analyse offener Daten

Unterschiedliche Einschätzungen liegen vor allem bei der Bewertung der Möglichkeiten der Integration und der Rolle einer Community innerhalb von ODAAaS vor. Experten, die ein verteiltes ODAAaS-Ökosystem als eine langfristige Perspektive ansehen, erkennen dabei weitreichendere Potentiale für die Integration einer Community. Generell unterscheiden sich die angedachten Konzepte bezüglich der Art von Akteuren, die sich in einem solchen System sinnvoll engagieren könnten. Hier reichen die möglichen Akteure von Entwicklern über Experten aus dem Bereich der Statistik und der statistischen Programmierung bis hin zu Domänenexperten der unterschiedlichen Datensätzen. Auch unterscheidet sich die mögliche Tiefe, in der solche Akteure in ODAAaS aktiv involviert werden können. Einige Fachleute sehen hier die Möglichkeit tiefgehender Integration und aktiver Intervention von Akteuren, z.B. im Bereich der Datenveredelung und der Erstellung von Hilfsmitteln. Andere sehen solche Akteure eher in einer passiveren Rolle als Wissensgemeinschaft, welche sich zum Beispiel über Wiki-Plattformen einbringen kann.

Bezüglich einer aktiven Beteiligung der Community bei der Generierung neuer offener Daten für Open Data-Portale liegen geteilte Meinungen vor. Ein Teil der Experten argumentiert, dass solche Daten nicht in eine Open Data-Plattform integriert werden sollten, um die klare Abgrenzung zwischen „offiziellen“ Daten (z.B. der öffentlichen Verwaltung) und daraus ableitbaren Datensätzen oder Teildatensätzen zu erhalten, da ansonsten Unklarheiten über die Provenienz der Daten entstehen könnten. Andere Akteure argumentieren, dass gerade solche Daten die Potentiale offener Daten noch weiter erhöhen können, da sich oft Fragestellungen von Nutzern nicht an der durch öffentliche Stellen veröffentlichten Form orientieren würden. Ein Interviewpartner argumentierte, dass solche grundlegenden Funktionen der Kombination von Daten bereits heute in Instrumenten der amtlichen Statistik Verwendung finden und dem Nutzer mehr Flexibilität erlauben. Solche Daten werden jedoch auch dabei nicht in den Bestand der offiziellen Indikatoren aufgenommen, so dass es sich hierbei eher um die Funktionalität zur Konstruktion von Ad hoc-Indikatoren handelt. Alle hierzu befragten Interviewpartner sehen jedoch im Kontext von ODAAaS Potentiale für Lösungen, die es Nutzern ermöglichen, Daten nach spezifischem Erkenntnisinteresse zu filtern, und sehen dies als notwendige Grundfunktionalität an.

Es ist aus Sicht der meisten befragten Experten davon abzuraten, nutzergenerierte Indikatoren in eine Open Data-Plattform Eingang finden zu lassen. Zumindest müsse ein entsprechender Mechanismus des Peer-Reviews implementiert werden.

6.2.7 Open Data Analytics – einfache vs. komplexe Konzepte

Hinsichtlich des Funktions- und Datenumfangs möglicher ODAAaS-Konzepte sehen einige Interviewpartner vor allem domänenspezifische Anwendungen mit klar abgegrenzten Nutzungskontexten im Vordergrund, was der in Kapitel 4.4 dargestellten Idee einer spezifischen Umsetzungskonzeption entspricht. Einige Experten argumentieren, dass ein ausschließlich auf Basis von Datenanalyse offener Daten basierender Markt nur schwer zu erschließen sei und an spezifischen Fragestellungen orientierte Lösungen sinnvoller seien. Der Vorteil liege dabei in einem klar abgrenzbaren Nutzungszweck. Der Betrieb eines generischen ODAAaS-Konzepts wird von diesen Interviewpartnern als momentan wenig sinnvoll eingestuft, da ständige Kommunikation und substantielles Commitment auf Seiten der Datenbereitsteller nicht vorausgesetzt werden kann oder entsprechende Pflegedienstleistungen nicht gewährleistet werden können. „One size fits all“-Konzepte erweisen sich laut einem Interviewpartner momentan in der Praxis als schwierig und sind ohne klar formulierte Anforderungen an die Analysemöglichkeiten kaum zu realisieren.

Andere Experten erkennen den Nutzen generischer ODAaaS-Konzepte, argumentieren aber, dass eine generische Lösung in Form einer Informationsplattform aus einer Hand den Aufwand der Umsetzung drastisch reduzieren könnte und Datenanalysefunktionen möglichst direkt in eine Open Data-Plattform integriert werden sollten.

Ein Teil der Interviewpartner sieht langfristig ein hohes Potential für offene schnittstellenbasierte Infrastrukturen. Offene Daten seien jedoch nur Teilmenge einer größeren Dateninfrastruktur. Hier sei insbesondere die verstärkte Etablierung und Umsetzung von Standards für verteilte offene Dateninfrastrukturen und eine flächendeckende Sicherstellung einer hohen Datenqualität notwendig. Des Weiteren argumentieren einige Experten, dass solche Lösungen nur dann langfristig nutzbringend seien, wenn sie mit entsprechenden einfachen und nutzerfreundlichen Werkzeugen verknüpft würden, die den Umgang mit heterogenen Diensten weitestgehend von technischem Vorwissen abstrahieren. Auch hier liegen jedoch kaum verbreitete technische Lösungen oder Ansätze vor, vielmehr sind diese momentan noch Gegenstand der aktuellen Forschung.

6.2.8 Abrechnungsmodelle für ODAaaS

Für AaaS-Lösungen werden unterschiedliche Abrechnungsmodelle angeboten. Je nach Anwendungsfall können nahezu alle gängigen Abrechnungsmodelle über eigene Lösungen oder Partnerlösungen realisiert werden. Welche Abrechnungsmodelle sinnvoll sind, sei insgesamt jedoch nur für den spezifischen Einzelfall zu klären. Nicht-kommerzielle Anbieter halten Modelle für vorstellbar, die die Nutzung des Dienstes kostenfrei ermöglichen, diese jedoch durch kostenpflichtige Beratungsdienste oder den Verkauf von Dokumentationen refinanzieren. Eine solche Variante sei jedoch stark von der intendierten Nutzergruppe und dem Anwendungsfall abhängig.

In integrierten Lösungen, die Dienste eines Anbieters in virtualisierter Form zur Verfügung stellen, sei ein Abonnentenmodell vorstellbar. Für Infrastrukturen, welche mehrere Dienstebenen integrieren, die nicht alle durch einen Anbieter vorgehalten werden können, stelle sich die Kalkulation von Preisen für Dienste als schwieriger dar, sei aber tendenziell über Abonnentenmodelle realisierbar. Für verteilte Dienstarchitekturen durch unterschiedliche Marktteilnehmer schlägt einer der Interviewpartner vor, dass ein Dienstportal mit integriertem Dienstkatalog ähnlich eines App-Stores sinnvoll wäre. Anbieter von Diensten sollen dabei einen prozentualen Teil der Einnahmen an den Betreiber des Dienstportals entrichten. Der Betreiber legt in dieser Variante die Art und Weise der Preiskalkulation fest, jedoch nicht den Preis selbst. Hier wird als Abrechnungsmodell ein „pay as you go“-Verfahren angeregt.

6.2.9 Apps für die Analyse offener Daten

Im Hinblick auf die Nutzung von Apps als Teil eines ODAaaS-Ökosystems sind aus Sicht der Experten unterschiedliche Varianten denkbar. Allgemein wird der Nutzen von Apps im Kontext Open Data als sehr hoch eingeschätzt. Dies gilt im Zusammenhang von AaaS vor allem für einfache Anwendungen und weniger für komplexe und offen gestaltete Systeme, in denen Nutzer eigene Funktionalitäten einbringen können. Einige Interviewpartner halten Apps aus dem BI-Bereich vor, welche zur Darstellung von Analyseergebnissen von zuvor klar abgegrenzten Analyseschritten genutzt werden. Auch bieten alle Anbieter entsprechende Schnittstellen an, welche von Apps genutzt werden können, um Ergebnisse an das (mobile) Endgerät zu übertragen. Die meisten Experten empfehlen, Smartphones nicht zur Berechnung von Ergebnissen heranzuziehen, sondern Smartphones als Darstellungsgerät zu verstehen. Allgemein wird der Einsatz von Apps in Form spezifischer Anwendungen mit klar begrenztem Umfang und einer klaren Zielgruppe als die sinnvollste Variante von ODAaaS-Apps dargestellt. Werden Apps im Kontext komplexer ODAaaS-Ökosysteme mit einem offenen Funktionsumfang eingesetzt, so

sehen die Interviewpartner vor allem einen Nutzen in der komplementären Anwendung von Apps zu umfangreicheren webbasierten Anwendungen.

6.2.10 Spezifische Anforderungen für ODAaaS

Die Barrieren, welche die Interviewpartner für ODAaaS-Konzepte sehen, sind nur selten technischer Natur. Zwar liegen auf technischer Seite zum Teil Einschränkungen vor, in welcher Weise ein ODAaaS-Ökosystem zeitnah und mit begrenzten Ressourcen implementiert werden kann, jedoch zeigt sich auch, dass sich diese überwiegend auf konkrete Ausgestaltungsoptionen beziehen. Marktfähige AaaS-Lösungen erlauben bereits heute unterschiedliche Umsetzungsvarianten. AaaS-Konzepte sind bereits auf dem Markt vorhanden, werden im industriellen Sektor schon heute verstärkt eingesetzt und sind technisch einfach übertragbar. Notwendige Voraussetzung für alle Konzepte ist die konsequente Nutzung offener Schnittstellen und offener Datenformate, welche über Open Data-Portale zum Teil gewährleistet sind.

Anforderungen werden überwiegend in nicht-technischen Aspekten gesehen und beziehen sich in erster Linie auf die zur Verfügung stehenden offenen Daten und die im internationalen Vergleich bisher in Deutschland noch wenig wahrgenommenen wirtschaftlichen Verwertungspotentiale offener Daten. Als Hemmnis werden hier die Umsetzungskosten und der verhaltene Investitionswille für die Entwicklung von generischen und auf offene Daten ausgerichteten Analyselösungen gesehen. Dies wird von einigen der Befragten darauf zurückgeführt, dass der Vorteil solcher nachgelagerten Analyseoptionen zum Teil nicht erkannt wird, bzw. der Nutzen vorwiegend in auf offenen Daten aufbauenden, spezifisch an einem konkreten Nutzungszusammenhang orientierten Lösungen gesehen wird. Avancierte Konzepte bedürfen laut Aussagen einiger Experten konkreter Beispiele, um den Vorteil der statistischen Analyse offener Daten zu verdeutlichen. Leichtgewichtige Anwendungen seien dabei eher von einfach erklärbarem und konkretem Nutzen und stellen damit leichter verständliche Beispiele für den Nutzen offener Daten dar. Gleichzeitig wird jedoch angemerkt, dass der Aufbau einer entsprechenden Infrastruktur hohe Kosten verursachen würde.

7 Varianten der Analyse offener Daten in der Cloud

Im Folgenden werden mögliche Ausgestaltungsformen für ODAaaS beschrieben. Hierbei werden die Ansätze dargestellt, welche bereits heute in Teilen umsetzbar sind. Bezüglich der Bereitstellung von AaaS als Cloud-Dienst lassen sich auf Basis der Interviews keine dedizierten Anforderungen ableiten. Für Analysewerkzeuge ist es danach technisch unerheblich, ob es sich um offene Daten handelt oder nicht. Wichtig ist die ausreichende Qualität der Daten. Grundsätzlich können nach Aussage der befragten Anbieter bei entsprechender Aufbereitung der Daten alle am Markt geläufigen AaaS-Werkzeuge genutzt werden. Durch die Verwendung standardisierter und offener Formate und Schnittstellen in Open Data-Portalen ist dies überwiegend der Fall. Etablierte kommerzielle Anbieter sowie eine Vielzahl kleinerer Anbieter halten bereits heute entsprechende Lösungen bereit.

Die Integration von Analysefunktionen in Open Data-Plattformen stellt einen Sonderfall dar, der jedoch langfristig eher als komplementär zu der Trennung von Open Data-Plattformen und AaaS-Lösungen gesehen werden kann. Im Allgemeinen werden Open Data-Plattformen einen Überblick über die Gesamtheit der vorhandenen Daten geben und ggf. einfache Visualisierungsunterstützung anbieten. Analysedienste werden dagegen auf die Beantwortung spezieller Fragestellungen zugeschnittene Lösungen anbieten. Zwar können ODAaaS-Konzepte, die auf der Integration von Open Data-Plattformen und Analysefunktionen beruhen, im Einzelfall sinnvoll sein, jedoch wird dadurch nur ein kleiner Ausschnitt möglicher Ausgestaltungsformen adressiert.

Während die befragten Anbieter zumeist Lösungen aus einer Hand bevorzugen, erlaubt das im Kapitel 4 beschriebene Modell die Kombination aufgabenbezogener Komponenten verschiedener Anbieter. Die Trennung der grundlegenden Schritte Datenbereitstellung, Datenbereinigung, Datenintegration und Datenanalyse kann dabei eine breitere Nutzung von Analysewerkzeuge unterschiedlicher Anbieter erlauben, welche zum Teil stark avancierte, statistische Verfahren umfassen. Aus Nachfragesicht kann eine Trennung dieser Schritte sinnvoll sein, da sie es dem Nutzer erlaubt, auf seine speziellen Erfordernisse zugeschnittene Lösungen einzusetzen. Erfahrenen Nutzern steht dabei ein größerer Funktionsumfang zur Verfügung. Dabei müssen jedoch die höheren Anforderungen an den Harmonisierungsgrad zwischen den einzelnen Funktionen eines ODAaaS-Systems beachtet werden.

Im Kapitel 3.1 wurden für den exemplarischen Aufbau von AaaS die allgemeinen Cloud-Konzepte bereits zu den Anforderungen für AaaS in Beziehung gesetzt. Hieraus ergibt sich die Frage, wie AaaS in einer Cloud-Umgebung umgesetzt werden kann. Zur Beantwortung dieser Frage lässt sich ergänzend das von der „Open Data Center Alliance“ entwickelte Nutzungsmodell (Open Data Center Alliance, 2013) heranziehen. Dort wird der Begriff Information as a Service definiert als *„The ability to provide standardized and secure methods to create, manage, exchange, and extract meaningful information from all available data in the right format at the right time.“*

Während der Schwerpunkt der Information as a Service-Definition auf der Befähigung zum Angebot von Analysewerkzeugen liegt (PaaS) und den Dienstanbieter adressiert, fokussiert die AaaS auf deren Nutzung (SaaS) als Endnutzer. Lassen sich die für die Datenanalyse angebotenen Werkzeuge in eine der Referenzarchitektur genügende Cloud-Infrastruktur einbetten, so liefert diese bereits Schnittstellen zu Sicherheitsfunktionen, Abrechnungsmechanismen und zu anderen Cloud-Systemen, wie sie beispielsweise zur Bereitstellung der Daten benötigt werden. Die Etablierung spezieller Lösungen für ODAaaS ist nicht mehr erforderlich. Abbildung 6 stellt beispielhaft eine Integration von AaaS-Werkzeugen in die Cloud-Referenzarchitektur von NIST dar. Ausgehend von der in Abbildung 3 dargestellten Aufteilung eines Cloud-Systems in verschiedene Dienstbereiche werden (O)DAaaS als spezielle SaaS-Ausprägung, Information as a Service als spezielle PaaS-Ausprägung und Storage as a Service als spezielle IaaS-Ausprägung dargestellt. Für ein ODAaaS-System ist allgemein unterstützende Software für das Management von Daten und Metadaten erforderlich. Zusammen mit der Verwaltung physikalischer

Ressourcen bilden diese Dienste die Grundlage für eine ODAaaS-Orchestrierung nach NIST, d.h. für die Kombination aufgabenbezogener, fachspezifischer Dienstbausteine. Vervollständigt wird das ODAaaS-System durch auf den Anwendungsfall zugeschnittene Management- und Sicherheitsdienste. Über die NIST-Rolle eines Brokers kann vermittelt auf Rohdaten oder vorausgewertete Daten zugegriffen werden, die ihrerseits als Data as a Service Cloud-Dienst angeboten werden.

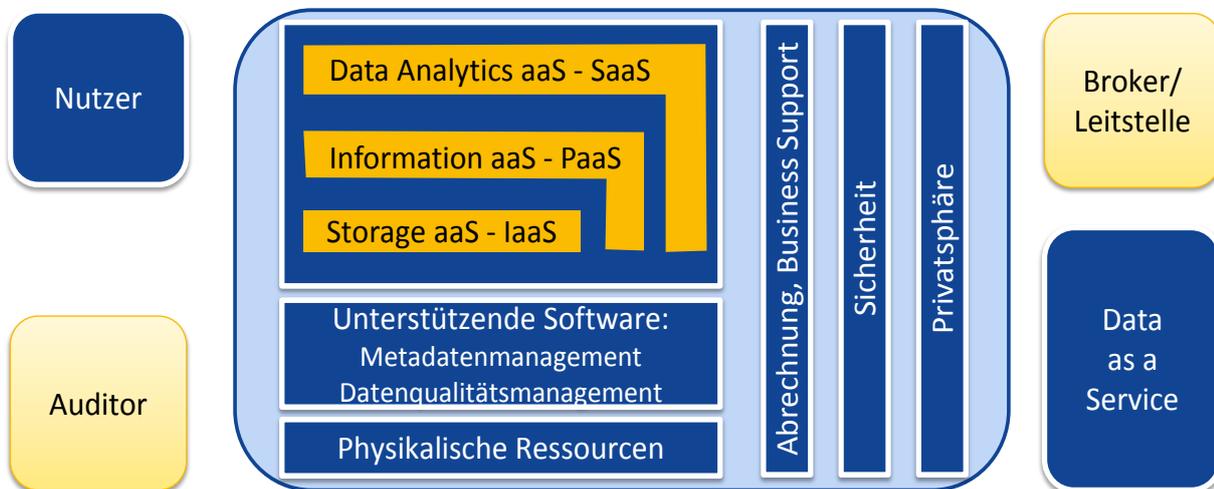


Abbildung 6: (O)DAaaS in der Cloud-Referenzarchitektur

Für Analysewerkzeuge ist demzufolge zu unterscheiden, ob das Werkzeug

- in einer Cloud-Infrastruktur betrieben und seinen Nutzern als Cloud-Dienst angeboten wird. In diesem Fall wird man von einem mandantenfähigen Dienst ausgehen, der den fünf Cloud-Charakteristiken genügt. Die zu analysierenden (offenen) Daten können entweder in der gleichen Cloud-Infrastruktur über Storage as a Service gespeichert, extern gespeichert oder extern als Data as a Service-Dienst angeboten werden. Für offene Daten wird in den meisten Fällen die zweite Variante zutreffen. Der AaaS-Anbieter stellt dem Kunden seine Dienstleistung entsprechend seinem Geschäftsmodell und den Lizenzbedingungen für die Nutzung der offenen Daten zur Verfügung.
- in einer Cloud-Infrastruktur über einer Information-/Integrationsplattform als Cloud-Dienst angeboten wird. In diesem Fall ändert sich für den Endnutzer nichts, der AaaS-Anbieter ist jedoch seinerseits Kunde eines PaaS-Dienstes. Für den Information as a Service-Anbieter ergeben sich dabei interessante Geschäftsmodelle, die von den Lizenzbedingungen der Daten unabhängig sind.
- ohne Cloud-Infrastruktur als extern gehosteter Dienst mandantenfähig betrieben und über standardisierte Internet-Protokolle angesprochen werden kann. Diese Variante kann nur bedingt als Cloud-konform betrachtet werden und entspricht eher einer Virtualisierung des Dienstes.

Abbildung 6 fasst die verschiedenen Möglichkeiten, die für die Durchführung einer Datenanalyse erforderlichen Schritte auf Cloud-Dienste abzubilden, zusammen. Dem an der Datenanalyse interessierten Endnutzer wird ein Software-Dienst angeboten, der die Auswertung und Darstellung der Daten ermöglicht. Optional besteht die Möglichkeit, die auszuwertenden Daten zu filtern, mit themenspezifischen Begriffen zu annotieren, in ein gemeinsames Format zu transformieren und letztendlich Daten aus verschiedenen Quellen zusammenzuführen. Diese Aufbereitung (Revise, Integrate) der Daten muss nicht zwangsweise von Endnutzer vorgenommen werden. An dieser Stelle sind Ansätze wie die Einbeziehung externer Nutzergruppen über *Crowdsourcing-Mechanismen* oder *maschinelles Lernen* vorstellbar, die eine Vorverarbeitung der Daten durchführen. Die zu Informationen aufbereiteten Daten werden dann in möglichst standardisierten Formaten gespeichert, wozu, insbesondere bei größeren Datenmengen, die Bereitstellung von Speicherkapazitäten über Cloud-Dienste sinnvoll ist.

Abbildung 7 beschreibt diese logische Referenzarchitektur für (O)DAaaS. Dabei werden den in Abbildung 6 eingeführten Ausprägungen der Cloud-Dienstmodelle die fachlichen Aufgaben des jeweiligen Modells und der typischen Nutzertyp des Cloud-Dienstes zugeordnet. Je stärker Analyse-Werkzeuge diese Architektur implementieren, desto besser ist deren Bereitstellung als Cloud-Dienst möglich.

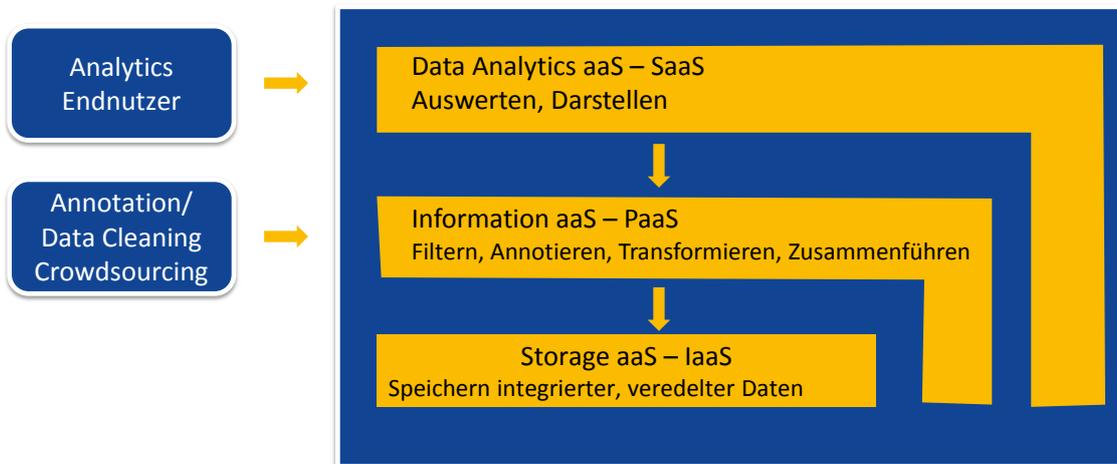


Abbildung 7: Datenanalyse durch kombinierte Cloud-Dienste

(O)DAaaS kann prinzipiell in verschiedenen Formen durch Cloud-Dienste implementiert werden. Ausgehend von einem monolithischen Ansatz, der ein Analysewerkzeug in einer virtualisierten Umgebung als Software-Dienst bereitstellt über einen hierarchisch strukturierten Ansatz, der ein Analysewerkzeug mittels IaaS und PaaS als SaaS bereitstellt, bis hin zu einem föderierten Ansatz, der eine verteilte Wertschöpfungskette implementiert, sind unterschiedliche Formen der Bereitstellung möglich.

Die Einschaltung von Datenvermittlern erlaubt den Aufbau von Wertschöpfungsketten auf Basis von Datenmarktplätzen. Unter Nutzung vermittelbarer, standardisierter Schnittstellen kann auf zu Informationen aufbereitete Daten zugegriffen werden und diese im Zuge einer weiteren Veredlung zu höherwertigen Informationen und letztendlich zu Wissen und Handlungsempfehlungen aufbereitet werden.

AaaS kann als kommerzielles Dienstangebot in einer Public Cloud angeboten werden. Ebenso ist es vorstellbar, dass für verwaltungsinterne Zwecke eine AaaS-Infrastruktur innerhalb einer, z.B. von öffentlichen IT-Dienstleistern betriebenen, Private- oder Community-Cloud betrieben wird. Betrachtet man den föderierten Ansatz, so ist auch eine kombinierte Hybrid-Cloud für die Bereitstellung der einzelnen Bestandteile des Gesamtangebots vorstellbar.

Ausgehend vom momentanen Stand der Technik und den Ergebnissen der durchgeführten Befragung ist somit davon auszugehen, dass zunächst die erstgenannten integrierten Ansätze realistisch und zeitnah umsetzbar sind, um ein erstes tragfähiges ODAaaS-System zu etablieren. Langfristig sollten jedoch auch hier die Potentiale eines Dienste- bzw. Datenmarktplatzes erschlossen werden.

8 Handlungsempfehlungen

Aus den im Rahmen dieser Studie gewonnenen Erkenntnissen lassen sich Handlungsempfehlungen für die beteiligten Akteure ableiten, die die Grundvoraussetzungen für die Einführung von ODAaaS verbessern und die Nutzung von ODAaaS erleichtern helfen. Hierbei werden sowohl technische als auch nicht-technische Empfehlungen gegeben.

8.1 Maßnahmen bei der Bereitstellung offener Daten

Etablierung einer Qualitätskultur für offene Daten

Datenqualität steht am Anfang jeder datenorientierten Arbeitsweise. Jedoch hat im Bereich Open Data noch keine flächendeckende Sensibilisierung für dieses Thema stattgefunden. Ein langfristiges Ziel im Rahmen der Etablierung von ODAaaS muss daher das Schaffen eines Bewusstseins für den Wert offener Daten in der öffentlichen Verwaltung sein. Hierzu soll der Nutzen qualitativ hochwertiger Daten verdeutlicht werden. Denkbar ist es, die zuständigen Akteure der öffentlichen Verwaltung durch Workshops und Leitfäden von der Wichtigkeit hoher Qualität bei der Veröffentlichung und Pflege offener Daten zu überzeugen.

Verbesserung der Qualität der beschreibenden Metadaten

Der Nutzen und die Analysierbarkeit offener Daten steigt beträchtlich durch die Vernetzung von Datenbeständen an. Für eine automatisierte Zusammenführung von Daten sind jedoch vor allem auch Metadaten unterhalb der Datensatzebene von Bedeutung. Die Zusammenführung von Daten kann umso einfacher gestaltet werden, je höher die Qualität der zu den Daten gehörenden Metadaten ist. Es ist anzustreben, dass in Open Data-Portalen weiterführende Metadaten hinterlegt werden, die einen stärkeren Sachdatenbezug aufweisen und über reine katalogartige Metadaten hinausgehen. Diese sollten über standardisierte Schnittstellen nachgelagerten Analysewerkzeuge zur Verfügung gestellt werden können.

Mit Linked Open Data liegen dabei bereits heute Konzepte vor, welche diesem Umstand Rechnung tragen. Jedoch liegt bislang nur ein sehr geringer Anteil offener Daten als Linked Open Data vor. Sollen die Potentiale von ODAaaS möglichst umfassend erschlossen werden, ist die Veröffentlichung von Daten als Linked Open Data weiter voranzutreiben. Dazu ist die öffentliche Verwaltung möglichst zeitnah mit dem Zugang zu entsprechenden Werkzeugen auszustatten. Mitarbeiter der öffentlichen Verwaltung sind im Umgang mit diesen Werkzeugen zu schulen. Zu den in diesem Kontext sinnvollen Maßnahmen gehört auch der Aufbau domänenspezifischer Ontologien, die sich an den Datenbeständen der öffentlichen Verwaltung und zugehörigem fachspezifischen Domänenwissen orientieren.

Unterstützung der öffentlichen Verwaltung durch Open Data Publisher Werkzeuge

Für die Unterstützung der Veröffentlichung offener Daten in benötigter Qualität sind geeignete Werkzeuge bereitzustellen. Hier besteht ein deutliches Ausbaupotential. Die Werkzeuge müssen zudem den Aufwand für die Bereitstellung offener Daten deutlich reduzieren. Denkbar sind hierbei vor allem intelligente Werkzeuge, die eine Validierung offener Daten zulassen und es erlauben Daten in semantisch ausdrucksstarker Form, z.B. direkt als Linked Open Data, zur Verfügung zu stellen.

Abbau von Medienbrüchen in der öffentlichen Verwaltung

Ein zentrales Hindernis für die Bereitstellung offener Daten besteht in dem Aufwand, maschinenbearbeitbare Repräsentationen mit hoher Datenqualität zu erzeugen und die Daten möglichst umfangreich mit Metadaten anzureichern. Der damit verbundene Aufwand wird derzeit häufig durch eine Vielzahl von Medienbrüchen negativ beeinflusst. Ein Abbau dieser Medienbrüche im Rahmen der Verwaltungsmodernisierung würde sich direkt auf die Qualität der offengelegten Daten und damit auf die Etablierung von ODAaaS auswirken.

Open Data Beauftragte in der öffentlichen Verwaltung etablieren

Die Bereitstellung offener Daten in automatisiert maschinenverarbeitbarer Form und deren verwaltungsübergreifende Analyse stellen für die öffentliche Verwaltung nicht zu unterschätzende Anforderungen dar. Indem Prozesse etabliert und Verantwortlichkeiten innerhalb der öffentlichen Verwaltung festgelegt werden, können gebündelte Kompetenzen innerhalb der öffentlichen Verwaltung aufgebaut werden. Die Benennung von Open Data-Beauftragten kann dabei ein wesentlicher Schritt sein, um die Bereitstellung und Analyse offener Daten effizient zu gestalten. Durch gezielte Schulungsaktivitäten zu Datenqualität, Analysemethoden und Potentialen statistischer Analyse können für die öffentliche Verwaltung Potentiale erschlossen werden, die eine Nutzung offener Daten zur Unterstützung strategischer Entscheidungen ermöglichen.

Veröffentlichung von Echtzeitdaten und Sensordaten als Open Data ausbauen

Echtzeit- und Sensordaten bieten im Gegensatz zu statischen Daten wesentlich umfangreichere Möglichkeiten, um interessante Anwendungskontexte zu erschließen. Solche Daten liegen derzeit jedoch nur sehr vereinzelt als Open Data vor. Beispiele sind Echtzeitdaten im Bereich des öffentlichen Nahverkehrs, meteorologische Daten oder z.B. Daten mit Bezug auf aktuelle Wartezeiten in Bürgerämtern. Hier bieten sich Möglichkeiten, die von einfachen Produkten bis hin zu umfangreichen Katastrophenwarn- und Informationssystemen reichen können.

Einbeziehung privater und offener Daten jenseits von Open Government Data

Open Data wird im politischen Diskurs oft auf Open Government Data verkürzt. Die bisher im deutschsprachigen Raum umgesetzten Open Data-Portale orientieren sich überwiegend an diesem Konzept, so dass wesentliche Potentiale der Analyse offener Daten nicht berücksichtigt werden. Der Anteil an offen zur Verfügung stehenden Daten geht jedoch über das hinaus, was unter Open Government Data verstanden wird. Die Einbeziehung privater und frei verfügbarer Daten aus Bereichen wie Forschung und Wissenschaft führt zu einer Steigerung des Nutzens für Analyse Zwecke. Entsprechende Sondierungsaktivitäten in diesem Bereich sind dringend anzuraten.

8.2 Maßnahmen bei der Analyse offener Daten

ODAAaaS als Entscheidungshilfe für die öffentliche Verwaltung aufbauen

Die Analyse offener Daten bietet nicht nur für die Information von Bürgern und für die Realisierung neuer Geschäftsmodelle interessante Optionen. Sie bietet auch für die öffentliche Verwaltung neue Impulse zur Generierung von Wissen und zur Unterstützung von Entscheidungsprozessen. Durch den Einsatz einfacher Analysefunktionen kann sowohl für interne Zwecke als auch im Dialog mit dem Bürger eine höhere Transparenz von Verwaltungsentscheidungen erreicht werden. So lassen sich durch den Einsatz von Datenanalyse und vor allem der Visualisierung offener Daten Inhalte zeitnah in Bezug auf konkrete Fragestellungen darstellen und vermitteln. Für interne Zwecke können Daten schnell und effizient abgefragt und mit Analysewerkzeugen bearbeitet werden. Generell steigt der Nutzen solcher Möglichkeiten mit dem Anteil an offengelegter Information. Je mehr Daten durch eine Verwaltung veröffentlicht werden, desto eher kann im Rahmen von ODAaaS mit offenen Daten gearbeitet werden und entsprechende Potentiale realisiert werden. Hierzu sind im Einzelfall

Schulungsmaßnahmen und Informationsveranstaltungen notwendig, die konkret an Datenbeständen und Fragestellungen der öffentlichen Verwaltung orientiert sind.

Verbesserte Bürgerinformationssysteme durch Datenanalyse

Weitere Möglichkeiten reichen hier von der verstärkten Visualisierung von Daten bis hin zu den Analysemöglichkeiten integrierter Daten. Hierdurch werden offene Daten auch für solche Bürger interessant, die einen einfachen Zugang zu objektiven Informationen suchen aber nicht über entsprechende Fachkenntnisse verfügen.

Demonstratoren für innovative, analysebasierte Apps bereitstellen

Heute werden bereits Apps als etablierte Möglichkeit genutzt, um offene Daten in einfacher Weise darzustellen. Dabei stehen momentan noch relativ einfache Informationssysteme im Vordergrund. Die Möglichkeiten, die sich aus der statistischen Analyse offener Daten in Kombination mit privaten Daten ergeben, sind dabei wesentlich vielfältiger anzusehen. Durch die Bereitstellung beispielhafter Analysewerkzeuge können weitergehende Möglichkeiten zur Bereinigung, Integration und Analyse von Daten aufgezeigt werden, die die Entwicklung innovativer Apps fördern.

Best Practice Beispiele der amtlichen Statistik

Mitarbeiter der amtlichen Statistik wie statistische Landes- und Bundesämter haben bereits heute einen umfangreichen Erfahrungsschatz im Umgang mit Daten und deren statistischer Auswertung. Konzepte wie das Fernrechnen, d.h. die Berechnung innerhalb einer Infrastruktur der statistischen Ämter auf Basis von nutzergenerierten Analyseskripten auf nicht öffentlich zugängliche Rohdatensätzen ohne Einsicht der Nutzer auf die Daten selbst, sind seit Jahrzehnten etabliert. Hier liegt eine Situation vor, die der Trennung von Daten und Metadaten bei Open Data Plattformen entspricht. Nutzer haben lediglich expliziten Zugriff auf Metadaten wie die Kodierung und Skalierung von Variablen.

Statistische Ämter haben eine lange Tradition in der Bereitstellung von Datenportalen, die über Eigenschaften einer Analyseplattform in Bezug auf Filterung, Kombination und Darstellung von Daten verfügen. Der wesentliche Unterschied zwischen heutigen Open Data-Portalen und der Veröffentlichung von Daten statistischer Ämter liegt in der genutzten IKT-Infrastruktur. In statistischen Ämtern sind Web-Plattformen direkt an Datenbanken angebunden, aus denen die zur Verfügung gestellten Daten bezogen werden. Die Daten unterliegen der unmittelbaren Kontrolle der statistischen Ämter. Die Bereitstellung von Daten und Metadaten in hoher Qualität ist eine der Kernkompetenzen statistischer Ämter. Dies kann als eine Blaupause für Umsetzungskonzepte innerhalb anderer Bereichen der öffentlichen Verwaltung dienen, bei denen Daten direkt aus der Infrastruktur der öffentlichen Verwaltung bezogen und web-basiert veröffentlicht werden. Das innerhalb der Verwaltungen vorliegende Wissen stellt eine gute Grundlage für die Etablierung eines Prozessmodells zur Bereitstellung und Analyse offener Daten dar. Dabei können auch Fragen, welche Art von Metadaten im Rahmen einer automatisierten Maschinenlesbarkeit und nachgelagerten Analyse sinnvoll sind, aus der Praxis erschlossen und übertragen werden. Die Identifikation und Übertragung von Best Practice-Beispielen aus statistischen Ämtern ist daher anzuraten.

8.3 Empfehlungen für Anbieter von AaaS-Werkzeugen

Bereitstellung offener Importschnittstellen

Die Entwicklung spezieller Analysewerkzeuge zur Analyse offener Daten ist aus wirtschaftlichen Gesichtspunkten wenig rentabel. Vielmehr ist die Bereitstellung offener Importschnittstellen anzuraten, über die ein Zugriff auf

gängige Formate offener Daten und Metadaten über Datenkataloge möglich ist. Der Import privater Daten muss ebenfalls unterstützt werden.

Bereitstellung von Funktionen zur Bereinigung und Integration von Daten

AaaS-Werkzeuge müssen die Bereinigung und Integration von Daten aus unterschiedlichen Quellen unterstützen. Dabei müssen die im Umfeld offener Daten vorhandenen Metdatenformate unterstützt werden.

Bereitstellung von AaaS-Diensten

Sowohl öffentliche Verwaltungen als auch Bürger werden in der Regel keine aufwendige Software zur Analyse offener Daten lokal installieren und betreiben. Vielmehr wird diese Funktionalität über Web-Schnittstellen mandantenfähig bereitgestellt und genutzt. Dabei konkurrieren allgemeine Analysewerkzeuge mit spezialisierten Apps. Werden offene und private Daten gemeinsam analysiert, so ist der Schutz der privaten Daten sicherzustellen. Gegebenenfalls müssen die Analysedienste dazu in öffentlichen Rechenzentren betrieben werden. Die Einhaltung rechtlicher Rahmenbedingungen ist zu beachten.

Dynamische Integration spezifischer Analyseverfahren

Für erfahrene Nutzer oder fachspezifische Erweiterungen ist die Integration spezifischer Analyseverfahren erforderlich. Die angebotenen Lösungen müssen daher PaaS-Charakteristika besitzen, die die dynamische Bereitstellung derartiger Verfahren ermöglichen.

Föderierte ODAaaS-Lösungen

Föderierte ODAaaS-Lösungen erlauben hohe Flexibilität, erfordern jedoch das Vorhandensein abgestimmter technischer Schnittstellen zwischen allen Teilnehmern auf dem Datenmarktplatz. Der konzeptionell vorhandenen Flexibilität einer derartigen Lösung steht die hohe technische Komplexität gegenüber. Die Realisierung eines verteilten, föderierten ODAaaS-Marktplatzes kann daher als wenig wahrscheinlich angesehen werden. Die Unterstützung offener Import- und Export-Schnittstellen für den Umgang mit offenen, privaten und vorausgewerteten Daten ist jedoch zwingend erforderlich.

9 Anhang – Fragebogen

Mit Aufkommen der Open Data Bewegung und der zugrunde liegende Transparenzgesetzgebung ergibt sich auch immer stärker die Notwendigkeit, Informationen aus diesen offenen Daten zu gewinnen. Experten sehen dabei Chancen für Politik, Wirtschaft und Zivilgesellschaft, sich weiter zu entfalten und das aus den offenen Daten gewonnene Wissen strategisch nutzbringend einzusetzen. Die Bereitstellung offener Daten erfolgt durch eine stetig wachsende Zahl von Portalen mit grafischen Nutzerschnittstellen und technischen APIs. Im Internet werden diverse Plattformen von Städten und Ländern angeboten, die nach und nach mit Daten gefüllt werden. Um Informationen aus den Daten zu gewinnen, müssen diese analysiert werden. Die dazu genutzte Software muss auf die angebotenen, technischen APIs zugreifen können und sollte möglichst, wie die Daten selber, ohne aufwendige Installationen nutzbar sein. Technisch bieten sich dazu Web-Anwendungen oder mobile Apps an, die zumindest im ersten Fall „as a Service“ in Cloud-Infrastrukturen betrieben werden können.

Ziel der Befragung ist es, im Rahmen einer für den ISPRAT e.V. erstellten Studie (Open Data Analytics as a Service – ODAaaS), von den Befragten angebotene Software-Lösungen auf ihre Eignung zur Auswertung offener Daten und einen Betrieb „as a Service“ zu untersuchen. Insbesondere sind wir an „Best Practice“-Lösungen interessiert, sofern die Befragten oder ihre Kunden bereits über praktische Erfahrungen in der Nutzung Ihrer Systeme zur Analyse offener Daten besitzen.

9.1 Technische Aspekte

9.1.1 Nutzerverwaltung und Barrierefreiheit

Verfügt Ihre Software über ein Administrationssystem z.B. über Nutzergruppen mit unterschiedlichen Rollen und Rechten?	
Sind unterschiedliche Rechte in Bezug auf die Datenhaltung und Datenbearbeitung möglich und in welcher Weise?	
W3C, IBM und ISO haben Vorgaben zu Barrierefreiheit erarbeitet. Sind Aspekte der Barrierefreiheit in Ihrer Lösung berücksichtigt und wenn ja in welcher Weise? (Augensteuerung, Mund-Maus, Sprachsteuerung,...)	
Welche Maßnahmen zur Sicherung der Interoperabilität, z.B. für unterschiedliche Browserversionen oder unterschiedliche Betriebssysteme sind in Ihrer Lösung berücksichtigt?	
Bietet Ihre Lösung die Möglichkeit der Kommunikation zwischen Nutzern bzw. kollaboratives Arbeiten und wie ist diese Lösung implementiert?	

9.1.2 Datenhaltung und -verarbeitung

Kann ihre Software Standards aus dem Bereich offener Daten interpretieren und auf durch Open Data Verzeichnisse über Metadaten referenzierte Daten	
--	--

zugreifen? Wenn ja, welche Standards werden unterstützt?	
Unterstützt Ihre Software Linked Open Data oder planen Sie, dies zu tun?	
Erlaubt Ihre Software das Zusammenführen von Daten aus unterschiedlichen Quellen?	
Welche Datenbanksysteme werden von Ihrer Lösung unterstützt?	
Besteht die Möglichkeit, verschiedene Datenbanken in das System einzubinden?	
Im Bereich des Open Data wird versucht, möglichst mit offenen Datentypen zu arbeiten. Welche Typen werden von Ihnen unterstützt? Wie gehen Sie mit speziellen Datenformaten, wie z.B. Geo-Daten um? Was halten Sie vom Import von nutzerspezifischen Datentypen, also eine Schnittstelle für usergenerierte Parser? <ul style="list-style-type: none"> - Import von Linked Data (RDF), JSON, XML, CSV --- GML,GPX,KML (Geoformate) - Export in RDF, XML, RDF, CSV, JP(E)G, PNG, RSS, HTML, GIF, EMF,EPS, SVG, GEOTIFF (Geoformat) - Metadatenstandards: SDMX,DDI, RDF 	
Unterstützt Ihr System auch die Analyse von Streaming Daten (z.B. Sensordaten)? In welcher Weise können diese durch Ihre Lösung in Echtzeit analysiert bzw. visualisiert werden?	
Open Data Plattformen bieten oft Daten in unterschiedlicher Qualität an. Manche Daten sind zurzeit nicht direkt als automatisiert maschinenlesbar einzustufen. Bietet Ihre Software Möglichkeiten diese Daten dennoch zu laden und zu bearbeiten und welche Werkzeuge (z.B. Daten-Cleaning) stehen hier zur Verfügung?	
Stellt Ihre Software Tools zur Datenselektion, z.B. dem Filtern von Daten bereit?	
Welche Arten von statistischen Analysemethoden sind in Ihrer Software integriert?	
Welche Visualisierungstechniken stellen Sie bereit?	
Besteht die Möglichkeit, Daten des Nutzers mit offenen Daten zu verbinden und durch welche Mechanismen und Werkzeuge wird dies in Ihrer Lösung realisiert? Welche Schutzmechanismen sieht die Software in diesem Fall vor?	
Können Sie sich Probleme bei der Nutzung von Open Data in Verbindung mit Closed Data vorstellen?	
Besitzt Ihre Software eine Funktion zur Versionierung der Daten, um z.B. gleiche Datensätze kollisionsfrei zu verarbeiten? Wenn, nein wie würden Sie sich ein solches System vorstellen?	

9.2 Datenschutz und Kontrollfunktionen.

Um die Nachvollziehbarkeit von Analysen zu erleichtern, werden auch Protokolle

und Historien benötigt. Welche Monitoring Tools stellt ihre Software entsprechend zur Verfügung?	
Das Urheberrecht ist im Falle der Datenanalyse ein relevanter Aspekt. Vor allem bei Integration und Abgleich offener Daten mit eigenen Daten. Sieht Ihre Lösung eine Signierung der neuen (Meta-)Daten vor?	
Wessen Aufgabe sollte es Ihrer Meinung nach sein, neue Informationen in Bezug auf Datenqualität und Urheber, zu zertifizieren? Halten Sie ein Zertifizierungssystem für sinnvoll? Zertifizierung durch die Community?	

9.3 Betrieb und Geschäftsmodelle

Welches Geschäftsmodell haben Sie oder empfehlen Sie für ODAAaaS?	
Verfügt Ihre Lösung über ein Abrechnungssystem und welche Varianten der Abrechnung werden hier angeboten? Was halten Sie von Mechanismen, wie z.B. von einem Angebot mit Schulungen für die Software, andere Einkommensquellen zu nutzen und die Software entsprechend kostenlos anzubieten?	
Welche Schnittstellen würden Sie für einen öffentlichen Zugang bereitstellen, um z.B. neue Informationen einfach zu erhalten oder auch verteilt zu bearbeiten? Besteht die Möglichkeit Ihre Analysewerkzeuge im Rahmen der Entwicklung und Nutzung von „Apps“ einzubinden?	
Ist die Software als SaaS einsetzbar? Kann die Software als Dienst oder als eigenes Kundenprodukt implementiert werden?	
Welche Voraussetzungen werden benötigt, um mit der Software umzugehen? <ul style="list-style-type: none"> - Expertensystem (vielseitiges Webtool) - Simple Klicksystem (Smartphone App) 	
Welche Nutzergruppen können Sie sich für die beiden Modelle vorstellen und welche bietet das höhere Potential? Welche Anwendungsfälle gibt es für die jeweiligen Nutzergruppen?	
Wie groß wäre der Aufwand einer generischen Lösung, bei der die Inhalte erst über Zeit aufgebaut werden, in Bezug auf Entwicklung und Einsetzbarkeit?	

9.4 Allgemeine Fragestellungen

Welchen Stellenwert nimmt Open Data in Ihrem Unternehmen ein?	
Wurde Ihre Software bereits zur Auswertung von Open Data genutzt? Welche Erfahrungen wurden dabei gemacht? Welche Probleme gab es? (teilnehmende	

Unternehmen, Themenbereiche)	
Sehen Sie einen Markt für „Analytics as a Service“- Geschäftsmodelle? Welche Barrieren könnte es geben/gibt es?	
<p>Wie würden Sie sich ein Open Data Analytics Ökosystem basierend auf Ihrer eigenen Technik vorstellen und wie würden Sie dabei Open Data Plattformen integrieren?</p> <ul style="list-style-type: none"> - Ökosystem: Angefangen bei der Datensammlung in den Open Data Plattformen über ein Black Box System mit Ihrer Technik bis zur Nutzung der neuen Informationen aus der Analyse 	
Wie hoch schätzen Sie den Aufwand zur Erstellung fachlicher Modelle, die die Grundlage für Auswertungen bilden, ein? Wer sollte derartige Modelle erstellen – Sie als Anbieter oder der Kunde selber?	
Wie stellen Sie sich die Datenhaltung und das Datenanalyse in so einem Ökosystem vor?	
In einem Open Data Analytics Ökosystem, wo sehen Sie die einzelnen Aufgaben in den Verwaltungen, den Open Data Plattformen und den Analytics Plattformbetreibern? Ist eine Community Beteiligung vorstellbar? Wenn ja, welche Aufgaben würde sie übernehmen? Wer übernimmt die Aufgabe des Cleaning?	
Können Sie sich die Plattform auch als einen Marktplatz für Dienstleistungen Dritter vorstellen, z.B. für Daten Cleaner oder Analysten?	
Welche Probleme sehen Sie beim Umgang mit Metadaten, z.B. in Bezug auf Datenqualität?	
Können Sie sich Potentiale und Hemmnisse Ihrer Software in der öffentlichen Verwaltung vorstellen? Welche Voraussetzungen müssten Ihrer Meinung nach in diesem Bereich bestehen?	
Haben Sie selbst noch bestimmte Anforderungen an eine Open Data Analytics Plattform?	
Wie sehen Sie den Übergang von klassischer Datenanalyse zu Big Data Analytics?	

10 Literaturverzeichnis

- BSI - Bundesamt für Sicherheit in der Informationstechnik. (Mai 2011). Eckpunktepapier - Mindestsicherheitsanforderungen an Cloud-Computing-Anbieter. (BSI, Hrsg.) Bonn.
- BSI - Bundesamt für Sicherheit in der Informationstechnik. (Mai 2011). *Eckpunktepapier-Mindestsicherheitsanforderungen an Cloud-Computing-Anbieter*. (BSI, Hrsg.) Abgerufen am November 2013 von <https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Mindestanforderungen/Eckpunktepapier-Sicherheitsempfehlungen-CloudComputing-Anbieter.pdf>
- CCUCDG - Cloud Computing Use Case Discussion Group. (July 2010). *Cloud Computing Use Cases - White Paper*. (cloudusecases.org, Hrsg.) Abgerufen am November 2013 von http://opencloudmanifesto.org/Cloud_Computing_Use_Cases_Whitepaper-4_0.pdf
- CSA - Cloud Security Alliance. (December 2009). *Security Guidance for Critical Areas of Focus in Cloud Computing*. Abgerufen am November 2013 von <https://cloudsecurityalliance.org/csaguide.pdf>
- DMTF - Distributed Management Task Force. (November 2009). *Interoperable Clouds - A White Paper from the Open Cloud Standards Incubator*. (Version 1.0.0). (DMTF, Hrsg.)
- DMTF - Distributed Management Task Force. (Januar 2010). *Open Virtualization Format (OVF)*. (DMTF) Abgerufen am November 2011 von <http://www.dmtf.org/standards/ovf>
- ISO/IEC JTC1/ SC38/WG3. (April 2013). *Cloud Computing Reference Architecture. Distributed application platforms and services — Cloud Computing — Reference Architecture*. (I. ISO/IEC, Hrsg.)
- Mell, P., & Grance, T. (July 2009). *The NIST Definition of Cloud Computing*. (Version 15).
- Mell, P., & Grance, T. (Januar 2011). *The NIST Definition of Cloud Computing (Draft)*. Abgerufen am November 2013 von http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf
- NIST - National Institute of Standards and Technology. (2010). *Cloud Computing Use Cases*. (NIST) Abgerufen am November 2011 von <http://www.nist.gov/itl/cloud/use-cases.cfm>
- NIST - National Institute of Standards and Technology. (2010). *Cloud System Use Cases (draft)*.
- NIST - National Institute of Standards and Technology. (2010). *Standards Acceleration to Jumpstart Adoption of Cloud Computing (SAJACC)*. (NIST) Abgerufen am November 2013 von <http://www.nist.gov/itl/cloud/sajacc.cfm>
- NIST - National Institute of Standards and Technology. (January 2011). *Cloud Architecture Reference Models: A Survey*. (NIST) Abgerufen am November 2013 von http://collaborate.nist.gov/twiki-cloud-computing/pub/CloudComputing/Meeting4AReferenceArchitecture013111/NIST_CCRATWG_004v2_Exist entReferenceModels_01182011.pdf
- NIST - National Institute of Standards and Technology. (September 2011). *NIST Cloud Computing Reference Architecture*. (NIST, Hrsg.) http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909505.
- NIST - National Institute of Standards and Technology. (2013). *Computer Security Resource Center - Publications*. (NIST) Abgerufen am November 2013 von <http://csrc.nist.gov/publications/PubsTC.html#Cloud%20Computing%20&%20Virtualization>

- OGF - Open Grid Forum. (May 2011). *Open Cloud Computing Interface - Core*. Abgerufen am November 2013 von <http://forge.ogf.org/sf/docman/do/listDocuments/projects.occi-wg/docman.root.specification>
- OGF - Open Grid Forum. (May 2011). *Open Cloud Computing Interface - Infrastructure*. Abgerufen am November 2013 von <http://forge.ogf.org/sf/docman/do/listDocuments/projects.occi-wg/docman.root.specification>
- Open Data Center Alliance. (März 2013). *Open Data Center Alliance*. Abgerufen am November 2013 von <http://www.opendatacenteralliance.org/>:
http://www.opendatacenteralliance.org/docs/ODCA_Commercial_Framework_MasterUM_v1.0_Mar2013.pdf
- Open Data Center Alliance. (2013). *Open Data Center Alliance*. Abgerufen am Oktober 2013 von <http://www.opendatacenteralliance.org/>:
http://www.opendatacenteralliance.org/docs/Information_as_a_Service_Master_Usage_Model_Rev1.0.pdf
- The Open Knowledge Foundation. (2013). *ckan - The open source data portal software*. (The Open Knowledge Foundation) Abgerufen am Oktober 2013 von <http://ckan.org/>
- vitako. (2013). *Bundes-Arbeitsgemeinschaft der Kommunalen IT-Dienstleister*. (vitako) Abgerufen am November 2013 von <http://www.vitako.de/>