XaaS und Big-Data-Technologien − cloudoptimiertes Datenmanagement für Open-Government-Data

Jens KRUMPE[1], Adrian KNOTH[2] und Burkhard GOLLA[3]

[1]egeo24.de · jens.krumpe@egeo24.de
[2]Friedrich-Schiller-Universität Jena
[3]Julius Kühn-Institut

Dieser Beitrag wurde durch das Programmkomitee als „reviewed paper" angenommen.

Einleitung

Bereits im Jahr 2011 schätzten die Analysten der International Data Corporation (IDC)[1] die verfügbare digitale Datenmenge auf 1 Milliarde Terabytes, welche bis zum Jahr 2020 um den Faktor 50 steigen soll. Nach der aktuellen IDC-Studie[2] werden in den kommenden Jahren mindestens 80 % des IT-Wirtschaftswachstums aus den Bereichen Mobile, Social, Cloud- und Big-Data generiert werden.

Vor dem Hintergrund der forcierten Harmonisierung und Bereitstellung von grenzüberschreitenden interoperablen und entgeltfreien Daten durch die europäische Gesetzgebung sowie Open-Data-Initiativen tragen die öffentliche Verwaltung und staatlich kontrollierte Stellen auf nationaler, europäischer und globaler Ebene zur Freisetzung von Massendaten (Big-Data) entscheidend bei.

Obwohl nutzbringende Informationsprodukte aus sogenannten Big-Data-Analysen durch die Unternehmen wie Google, Facebook oder Twitter allgegenwärtig sind, reagiert die Informationsbranche zögerlich auf die Erschließung neuer Geschäftsfelder auf Basis multithematischer und offener Verwaltungsdaten. Zum einen ist dieser Umstand auf die Heterogenität der Lizenzbestimmungen, der Semantik, der Syntax sowie uneinheitliche Qualitätsstandards von dezentral bereitgestellten Verwaltungsdaten zurückzuführen. Zur Stimulation der Geoinformationswirtschaft müssen zum anderen datenparallele Methoden zur Analyse von Massendaten mit Raumbezug entwickelt werden, um Synergien zwischen Cloud-Diensten und den Geowissenschaften herzustellen (YANG et al. 2011). Bei der Ableitung von nutzbringenden Informationsprodukten aus dezentral bereitgestellten Verwaltungsdaten gilt es weiterhin, stets die Authentizität und Integrität gegenüber den Originaldatensätzen zu wahren. Um das Potenzial von Open-Government-Data vollständig auszuschöpfen, sind Daten bereitstellende Verwaltungen als auch Daten konsumierende Institutionen (Wirtschaft, Verwaltung) deshalb gezwungen, ein verantwortungsbewusstes Datenmanagement zu betreiben.

[1] IDC-International Data Corporation (2012), http://tinyurl.com/a9gtaan/ (31.01.2013)
[2] IDC-International Data Corporation (2012), http://tinyurl.com/a5ctbcs/ (31.01.2013)

Strobl, J., Blaschke, T., Griesebner, G. & Zagel, B. (Hrsg.) (2013): Angewandte Geoinformatik 2013.
© Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach. ISBN 978-3-87907-533-1.
Dieser Beitrag ist ein Open-Access-Beitrag, der unter den Bedingungen und unter den Auflagen der Creative Commons Attribution Lizenz verteilt wird (http://creativecommons.org/licenses/by/3.0/).

Im Folgenden wird ein Konzept für ein prozessorientiertes Datenmanagement vorgestellt, welches vor dem Hintergrund der Ableitung eines Geoinformationsprodukts aus dezentral bereitgestellten Verwaltungsdaten die Potenziale und Herausforderungen von Cloud-Diensten (XaaS) aus dem Blickwinkel der Wirtschaft, Wissenschaft und Verwaltung beleuchtet. Die nutzbringenden Eigenschaften des vorgestellten Konzepts werden mit einem Diskussionspartner aus der Bundesverwaltung bewertet.

1 Motivation

Die Open-Government-Initiativen der EU, des Bundes und der Länder zielen darauf ab, das weitgehend ungenutzte Potenzial von Verwaltungsdaten für Gesellschaft und Wirtschaft zu heben. Die durch Infrastrukturen derzeit und zukünftig zugänglichen dezentral bereitgestellten Daten aus öffentlicher Verwaltung oder staatlich kontrollierten Stellen wachsen exponentiell und erfüllen mit der Vielfalt an strukturierten und unstrukturierten Daten alle Merkmale von Massendaten (engl. Big Data).

Die Bereitstellung und Nutzung dieser Massendaten erfolgt überwiegend nach dem SOA-Entwurfsmuster[3] und stellt nach SCHEPERS et al. (2008) damit hohe organisatorische und technische Ansprüche an die Datenbereitsteller und -Konsumenten, welche unter dem Begriff SOA-Governance zusammengefasst werden.

Bedingt durch die technikgetriebene Verwaltungsreform (E-Government) werden beispielsweise Datenkonsumenten in die Lage versetzt, Daten unterschiedlichster Thematiken medienbruchfrei in ihre Geschäftsprozesse zu integrieren und weiterzuverarbeiten. Im Ergebnis dieser Weiterverarbeitung entstehen neue Daten oder Entscheidungsgrundlagen, welche vor dem Hintergrund der Wahrung der Nachvollziehbarkeit, Beweiserhaltung und Referenzierbarkeit besonders behandelt werden müssen. Um Skaleneffekte aus offenen Verwaltungsdaten zu erreichen, empfiehlt sich eine ganzheitliche Betrachtungsweise der notwendigen rechtlichen, organisatorischen- und technischen Maßnahmen, welche nachfolgend als Datenmanagement bezeichnet werden. Echtzeitanalysen großer, unstrukturierter Daten zur Ableitung von Ad-hoc-Entscheidungsgrundlagen durch künstliche Intelligenz verstärken die Anforderung an ein effektives und effizientes Datenmanagement. Derzeit erfordert das Datenmanagement von Massendaten, die bspw. aus Datenfusionen entstehen können, größte Anstrengungen beim Datenkonsumenten: Die stetig wachsende Datenmenge zwingt ihn, technische und personelle Ressourcen zum Zweck des Datenmanagements wie Datenbereinigung, -restrukturierung, -kombination, -fusion bis hin zur nachnutzbaren Archivierung des abgeleiteten anwendungsspezifischen Informationsprodukts vorzuhalten und somit Mittel zu binden, die seinem eigentlichen Kerngeschäft entzogen werden.

Es besteht daher die Motivation, ein Lösungskonzept zu erarbeiten, welches unter Berücksichtigung von cloudoptimierten Dienstleistungsangeboten auf die Vereinfachung, Effizienzsteigerung und Kostenminimierung des Kerngeschäfts des Datenkonsumenten abzielt.

[3] INSPIRE Download Services, http://tinyurl.com/bpu4sba/ (31.01.2013)

2 XaaS und BigData-Technologien

Mit zunehmender Reife verfügbarer Cloud-Infrastrukturen steigt die Nachfrage nach XaaS (Anything as a Service). Wurde zu Beginn lediglich fremde Rechentechnik in Anspruch genommen (Infrastructure as a Service, IaaS), so dehnt Software as a Service (SaaS) dieses Miet- bzw. Mitnutzungskonzept auch auf die eingesetzte Software aus. Sollen große Datenmengen (Big Data) analysiert werden, bietet es sich an, auch die Datenbestände auf gemieteten Cloud-Komponenten vorzuhalten, um so flexibel auf wechselnde Speicheranforderungen reagieren zu können, ohne eine kostspielige eigene Infrastruktur betreiben zu müssen.

Big Data kennzeichnet, dass die Daten ob ihres Umfangs nicht mehr mit traditionellen Datenbanktechnologien verarbeitet werden können. Stattdessen kommen skalierbare Clustersysteme zum Einsatz, die je nach Bedarf dynamisch in ihrer Größe angepasst werden. Der Betrieb solcher Clustersysteme erfordert spezialisierte Rechenzentren, die dabei von Skaleneffekten und Fixkostendegression profitieren und damit für Endkunden wirtschaftlich günstiger sein können als eine eigene Infrastruktur.

Google bietet mit der Earth Engine[4] eine für die Umweltdomäne bereitgestellte Infrastruktur, bestehend aus Fernerkundungsdaten, in einer optimierten Datenhaltung und einem speziellen Programmiermodell für Big-Data-Analysen und kommt damit dem nachfolgend beschriebenen Konzept eines cloudoptimierten Datenmanagements am nächsten.

3 Konzept

Bei der Ableitung nutzbringender Informationsprodukte aus dezentral bereitgestellten Verwaltungsdaten durch Big-Data-Analysen empfiehlt es sich, Massendaten weitestgehend dort zu analysieren, wo sie erzeugt und gespeichert werden. Andernfalls würde bei der Datenübermittlung zwischen Datenspeicherungs- und Analysesystem ein Flaschenhals entstehen, der sich für zeitkritische Anwendungen verbietet. Um Mehrwerte aus Verwaltungsdaten zu generieren, muss ein adäquates Datenmanagement, vorzugsweise auf der gleichen technischen Plattform betrieben werden. Wird die technische Plattform durch eine cloudoptimierte Infrastruktur realisiert, sind organisatorische Verantwortungsbereiche zu schaffen. Durch die klare Trennung und Verlagerung des Datenmanagements auf die Verantwortungsbereiche der hierfür identifizierten Akteure *Datenbereitsteller* (Data-Provider), *Datenaufbereiter* (Cloud Service Provider) und *Plattformbetreiber* (Cloud Infrastructure Provider) kann eine Vereinfachung des Kerngeschäfts beim *Datenkonsumenten* (Enduser) erreicht werden. Wie in Abbildung 1 dargestellt wird, fällt in den Verantwortungsbereich des Datenaufbereiters das Datenmanagement von fach- bzw. anwendungsspezifischen Infrastrukturdaten sowie das Bereitstellen abgeleiteter Informationsprodukte in Form von Software-(SaaS), Daten-(DaaS) oder Analyse-Services (AaaS). Da offene Verwaltungsdaten eine Spezialisierung allgemeiner Forschungsdaten darstellen, sollen anschließend die im Curation-Life-Cycle-Modell[5] empfohlenen Phasen des Lebenszyklus von Forschungsdaten auf die für die Ableitung nutzbringender Informationsprodukte erarbeiteten Teilpro-

[4] Google Earth Engine (2012), http://earthengine.google.org (31.01.2013)
[5] DCC Curation Lifecycle Model (2010), http://preview.tinyurl.com/2699oob (31.01.2013)

zesse wie Vertragsmanagement und Datenselektion, Datenübermittlung und Qualitätssicherung, Datenspeicherung und Informationsproduktbereitstellung angewendet werden.

Die Bewertung der nutzbringenden Eigenschaften des vorgestellten Konzepts erfolgte am Beispiel eines Geschäftsprozesses innerhalb des Julius Kühn-Instituts (JKI), einer Ressorteinrichtung des Bundes. In diesem Geschäftsprozess werden Stoffflüsse in Kleinstgewässern auf Basis des in den landschaftsbeschreibenden Geodaten nachgewiesenen Gewässernetzes simuliert. Als Bundesforschungsinstitut berät das JKI die Bundesregierung. Für die Arbeit des JKI ist es essenziell, trotz des zu erwartenden Datenwachstums von zeitlich und räumlich immer höher aufgelösten Daten handlungsfähig zu bleiben. Vor dem Hintergrund durch INSPIRE bereitgestellter Datensätze erfordern Stoffflusssimulationen beispielsweise eine zusätzliche technische Ausstattung sowie spezielles Know-how im Umgang mit Massendaten. In die Bewertung fließt ausschließlich der vom JKI abgeschätzte Effizienzgewinn gegenüber dem derzeitigen Geschäftsprozess ein. Der Bewertungsmaßstab wurde mit den Merkmalen ‘moderat’, ‘hoch’ und ‘sehr hoch’ ordinal skaliert.

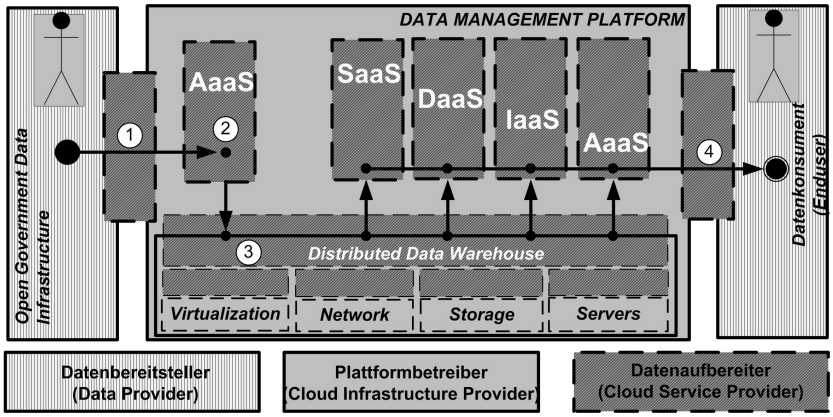

Abb. 1: Prozessorientiertes, cloudoptimiertes Datenmanagement

3.1 Vertragsmanagement und Datenselektion (Teilprozess 1)

Die subsidiäre Struktur Deutschlands und der EU sind die Voraussetzungen für die Selbstständigkeit der Verwaltungen. Demnach werden Infrastrukturdaten aus den Verwaltungen nicht zentral sondern dezentral bereitgestellt. Eine Ausnahme bilden die Open-Data-Portale[6], die einen zentralisierten Zugriff auf Daten unterschiedlicher Verwaltungen und Verwaltungsebenen ermöglichen. Dieser Teilprozess subsumiert die nach DCC(2010) empfohlenen Arbeitsschritte *Create* and *Receive*, *Appraise Select* und *Dispose* und hat zum Ziel, den notwendigen Dialog mit den einzelnen Datenbereitstellern zu zentralisieren und somit zu vereinfachen. Das Vertragsmanagement und die Datenselektion von anwendungsspezifischen Infrastrukturdaten gehören zum Geschäftsfeld des Datenaufbereiters und liegen damit in dessen Verantwortungsbereich. Landschaftsbeschreibende Geodaten im Maß-

[6] Berlin Open Data (2012), http://daten.berlin.de/ (31.01.2013)

stab bis 1:50000 werden bislang zu unterschiedlichen Konditionen gegenüber Wirtschaft und Verwaltung abgegeben, weshalb der Mehrwert eines zentralisierten Vertragsmanagements nicht erprobt werden konnte.

Aus Sicht der Verwaltung ist von einer *moderaten* nutzbringenden Eigenschaft auszugehen. Die für den Teilprozess notwendigen Aufgaben bezüglich der Wahl des Datenformats, der Datenstruktur, der Methode zur Qualitätssicherung, technische und lizenzrechtliche Abschätzungen der Datenverarbeitung und Strategieüberlegungen zur Archivierung von differenziellen oder vollumfänglichen Datenaktualisierungen werden vollständig durch den Datenaufbereiter realisiert.

3.2 Datenübermittlung und Qualitätssicherung (Teilprozess 2)

Dieser Teilprozess subsumiert die nach DCC(2010) empfohlenen Arbeitsschritte *Ingest* und *Preservation*. Nach der Übermittlung müssen die Daten auf deren Vollständigkeit überprüft und die Übermittlung dokumentiert werden. Des Weiteren sind Maßnahmen zur nachhaltigen Sicherung der Integrität, Authentizität und Nutzbarkeit der Daten zu ergreifen, sofern diese nicht bereits von der bereitstellenden Institution getroffen wurden. Im Anschluss erfolgt meistens die Dekompression, eine Formatumwandlung sowie die Überprüfung der Daten auf semantische und syntaktische Richtigkeit. Dieser Teilprozess ist für jeden Datensatz aus der Open-Data-Infrastruktur zu wiederholen. Die zentrale Qualitätssicherung von Infrastrukturdaten schlagen auch MOHAMMADI et al. (2010) als nutzbringende Eigenschaft für den Datenkonsumenten vor.

Aus Sicht der Verwaltung ist von einer *sehr hohen* nutzbringenden Eigenschaft auszugehen: Bereits das Überprüfen qualitativer Eigenschaften einer Objektart im ATKIS-DLM ist eine zeitaufwendige aber für die Extraktion von Informationen zur Entscheidungsunterstützung wichtige und notwendige Aufgabe. Beispielsweise wird im ATKIS-DLM das Knoten-Kanten-Modell der Fließgewässer[7] dokumentiert aber nicht explizit semantisch abgebildet. Jede Aktualisierung der Infrastrukturdaten erfordert das Regenieren einer expliziten Abbildung des Knoten-Kanten-Modells zum Zweck der Qualitätssicherung. Ein weiteres Beispiel ist, dass die attributive Eigenschaft der Gewässerkennzahl der Objektart Gewässerachse nach AdV(2009) mit 30 Zeichen zu codieren ist. Die von den Landesvermessungsämtern bereitgestellten ATKIS-DLM-Bestandsdatenauszüge weisen entgegen der Dokumentation in Sachsen-Anhalt[8] zwei und in Rheinland Pfalz[9] 20 Zeichen auf. Auch die Verwendung unterschiedlicher Versionen[10] des konzeptionellen ATKIS-Modells gibt einen Eindruck möglicher Fehlerquellen, die vor einer Datenanalyse zur Informationsextraktion bereinigt werden müssen. Vor dem Hintergrund der zukünftig stetig wachsenden und in der Struktur zunehmenden heterogenen Datenmengen, wird anhand des Beispiels die Bedeutung einer dienstleistungsorientierten, verlässlichen und performanten Qualitätssicherung durch einen Datenaufbereiter sichtbar. Neben der Qualitätssicherung binden Arbeitsschritte zur Datenübermittlung und Datenbereitstellung personelle und technische Ressourcen beim Datenkonsumenten. Die Datenübermittlung jedes einzelnen Infrastrukturdatensatzes wird bei-

---

7 ATKIS-OK Basis DLM 6.0 (2008), http://tinyurl.com/autpuvg (31.01.2013)
8 ATKIS Bestandsdatenauszug Sachsen Anhalt (2011), http://tinyurl.com/a5dfesm (31.01.2013)
9 ATKIS Bestandsdatenauszug Rheinland Pfalz (2009), http://tinyurl.com/czgc4yc (31.01.2013)
10 AAA-Umfrage zum Migrationsstand (2012), http://tinyurl.com/agdwm2c (31.01.2013)

spielsweise durch den verwaltungstechnischen Geschäftsprozess „Posteingang“ bearbeitet, er zusätzlich eine Verzögerung im Betriebsablauf hervorruft. Auch für Geschäftsprozesse wie Katalogisierung, Archivierung sowie Zwischenspeicherung von Daten für Qualitätskontrollroutinen müssen personelle Ressourcen sowie Hard- und Softwaresysteme bereitgestellt werden. Für den beschriebenen Teilprozess wurden beim Datenkonsumenten bislang drei Mann-Monate veranschlagt.

3.3 Datenspeicherung (Teilprozess 3)

Dieser Teilprozess entspricht dem Arbeitsschritt *Store* der DCC-Empfehlungen und speichert die qualitätsgesicherten Infrastrukturdaten in einer redundanten aber für Analysezwecke optimierten Datenstruktur. Durch das benötigte domänenspezifische Wissen fällt der Entwurf dieser optimierten multidimensionalen Datenstruktur, klassischerweise als Data-Warehouse bezeichnet, in den Verantwortungsbereich des Datenaufbereiters. Für Big-Data-Analysen eignen sich nach CAO et al. (2011) und KAUR et al. (2012) insbesondere cloudoptimierte Data-Warehouses (DW). Das Entwickeln von multidimensionalen Datenmodellen insbesondere für die raumbezogene Wissensextraktion ist derzeit noch ein aktives Forschungsfeld (MALINOWSKI 2011), (KYUNG et al. 2012).

Da nach der PSI-Richtlinie[11] Infrastrukturdaten einer standardisierten Open-Data-Lizenz unterliegen sollten, sind Bedenken hinsichtlich des Zugriffsschutzes irrelevant, für schützenswerte anwendungsspezifische Daten hingegen bieten Plattformbetreiber Datenverschlüsselungen[12] an.

Aus Sicht der Verwaltung ist von einer *hohen* nutzbringenden Eigenschaft auszugehen. Das Speichern der Daten in der Cloud bietet gegenüber der eigenen Infrastruktur die Vorteile der automatischen Datenreplikation, der elastischen Speichererweiterung für bspw. kontinuierliche Datenströme, der Energieeffizienz (Green-IT) sowie der Kostenreduktion für Hard- und Software und Wartungspersonal.

3.4 Informationsproduktbereitstellung (Teilprozess 4)

Dieser Teilprozess subsumiert die nach DCC (2010) empfohlenen Arbeitsschritte *Access, Use, Reuse* und *Transform*. Die Informationsproduktbereitstellung ist aus Sicht des Datenkonsumenten die einzige direkte Schnittstelle zum Datenaufbereiter. In diesem Teilprozess stellt die zielgerichtete Auffindbarkeit von Daten, Diensten und Informationsproduktbereitstellung das Hauptgeschäftsfeld des Datenaufbereiters dar.

Aus Sicht der Verwaltung ist von einer sehr hohen nutzbringenden Eigenschaft auszugehen. Der Zugriff und die Nutzung der Informationsprodukte können durch unterschiedliche Informationsschnittstellen wie Handy, eBook oder RFID und über verschiedene Servicemodelle (DaaS, AaaS, SaaS) bereitgestellt werden. In Verbindung mit dem ortsungebundenen Datenzugriff verbessert sich aus Sicht des Datenkonsumenten die verwaltungsübergreifende Zusammenarbeit. Unter Verwendung des DaaS-Modells können transformierte und qualitätsgesicherte domänen- oder anwendungsspezifische Daten, bspw. das o. g. Gewässernetz, in einem vom Datenkonsumenten präferierten Format zugänglich gemacht werden. Data-

[11] Directive 2003/98/EC, http://tinyurl.com/yg48s2o (31.01.2013)
[12] Amazon WS, Server Side Encryption Support, http://tinyurl.com/amuq9rf (31.01.2013)

Marts als Untermenge des DW aber auch dedizierte Datenmanagementaufgaben wie das Archivieren von nutzerspezifischen Analysedaten formen das Produktportfolio des Datenaufbereiters. DaaS kann sowohl in SaaS- als auch in AaaS-Betreiberkonzepte verwendet werden. Bei einem AaaS-Betreiberkonzept haben die Datenkonsumenten die Möglichkeit, SQL-ähnliche Ad-hoc-Abfragen an das DW durchzuführen. Das Betreiberkonzept SaaS sieht vor, dass der Datenaufbereiter vor dem Hintergrund nutzerspezifischer Daten und Anforderungen Software entwickelt und dem Datenkonsumenten durch die Cloud verfügbar macht (bspw. Stoffflusssimulationen). Die unter 3.2 aufgeführten Potenziale zur Kostenminimierung von technischen und personellen Infrastrukturen verstärken die nutzbringende Eigenschaft des hier vorgestellten Konzepts. Je nach Priorität der Aufgaben des Datenkonsumenten können diese in einer skalierbaren Infrastruktur (IaaS) abgearbeitet werden.

4 Ergebnisse, Chancen und Herausforderungen

Für die Implementierung der aufgeführten Teilprozesse muss eine zum Erreichen des Unternehmensziels geeignete Cloud-Architektur gewählt werden. Während offene Daten überwiegend in einer der letzten Phasen des Lebenszyklus von Verwaltungsdaten entstehen und sensible Daten in früheren Phasen analysiert oder verwaltet werden, kann die Nutzung hybrider Cloud-Architekturen empfohlen werden. Die technische Machbarkeit des vorgestellten Konzepts wurde unter Verwendung des Hadoop-Ökosystems (Hadoop, HBase, Hive und Mahout) des Plattformbetreibers Amazon belegt. Neben Amazon existieren eine Reihe alternativer Plattformbetreiber, bspw. Microsoft oder Hortonworks.

Das gezeigte Konzept ist vordergründig zur Effizienzsteigerung datenintensiver Geschäftsprozesse in der Verwaltung entwickelt worden und stellt damit einen konkreten Bezug zur E-Government-Strategie der Verwaltungen her. Da ein modernes E-Government die Kooperation von Wirtschaft, Wissenschaft und Verwaltung zwingend voraussetzt, wird im folgenden Abschnitt auf die wichtigsten Ergebnisse für diese drei Interessengruppen eingegangen.

4.1 Aus Sicht der Wirtschaft

Das Konzept stellt eine cloudoptimierte Datenmanagementplattform vor, in der u. a. Big-Data-Analysen mithilfe datenparalleler Methoden durchgeführt werden können. Die Verwendung der von der Open-Source-Community unterstützten Map-Reduce (MR) Programmierschnittstelle zur Entwicklung spezieller Softwarebausteine für Informations- und Wissensextraktion aus Massendaten birgt sowohl aus Kunden- als auch aus Dienstleistersicht gegenüber angepassten MR-Implementierungen wie StratoSphere PACT[13] oder Googles aktueller MR-Version (Jan/2013) eine Reihe von Vorteilen. Es besteht die Möglichkeit, auf den Wettbewerb unter den Plattformanbietern ohne Anpassung der Software flexibel zu reagieren. Neben den daraus resultierenden monetären Vorteilen können die Softwarebausteine bei besonders geschützten Daten problemlos in einer Private-Cloud (Behörden-Cloud) des Kunden betrieben werden, ohne dass eine softwareseitige Anpassung durch den Dienstleister erfolgen muss. Weiterhin profitiert der Kunde unmittelbar von der Innovationskraft der Open-Source-Community, welche die Weiterentwicklung ständig vorantreibt.

[13] StratoSphere (2012), https://stratosphere.eu/ (31.01.2013)

Aus den o. g. Architekturskizzen ist aus Sicht der Autoren ein hinsichtlich Verantwortlichkeiten und Kernkompetenzen der Akteure klar strukturierter Geschäftsprozess ableitbar, der die Teilprozesse aller Beteiligten vereinfacht und somit zur Leistungs- und Qualitätssteigerung in allen Verantwortungsbereichen beiträgt. Durch die oben beschriebene zentralisierte Dienstleistung (Single Point of Update) zur Bereitstellung von veredelten Geodaten oder extrahiertem Wissen mit XaaS-Betreibermodellen und Big-Data-Technologien entfallen die periodischen Aufgaben des Datenmanagements wie Datenbeschaffung, Qualitätsanalyse, Datenaufbereitung, Datenimport und Datenanalyse vollständig auf den Dienstleister.

Cloudoptimierte Infrastrukturen bieten den Vorteil, tatsächliche Nutzungen in Minutenauflösung abzurechnen. Es besteht zunehmend die Herausforderung, Marketingstrategien und für den Endkunden nachvollziehbare Qualitätskriterien für effizienten und somit kostensparenden Programmcode zu entwickeln.

Plattformanbieter wie Amazon nehmen Start-Ups der Big-Data-Branche das Risiko ab, in eigene IT-Infrastrukturen zu investieren. Die Fokussierung der Start-Ups auf die Geschäftsidee hilft, die Produkteinführungszeit „time to market“ zu minimieren. Innovative Lösungen für Gesellschaft und Verwaltung können damit schneller verfügbar gemacht werden.

Um Skaleneffekte aus offenen Verwaltungsdaten zu erzielen, müssen deren Datenmodelle und Austauschformate standardisiert werden. Die in Relation zu der Vielfalt an unterschiedlichen Daten aus den Verwaltungen wenig dokumentierten Standards (bspw. NAS) im XRepository[14] führt zu Unsicherheiten bei potenziellen Systemlösungsanbietern, da die Entwicklung eines rentablen Geschäftsmodells aufgrund schwer abschätzbarer Softwarekosten schwierig ist.

Neben den allgemeinen zu berücksichtigenden Thesen[15] zur Sicherheit und Datenschutzaspekten in der Cloud wird die Wertschöpfung durch Mehrwertdienste für Gesellschaft, Wirtschaft und Verwaltung aus den Open-Government-(Data)-Initiativen durch uneinheitliche Lizenzregelungen nach wie vor gebremst. Sowohl die für das Open-Data-Portal der Bundesregierung in Entwicklung befindlichen Lizenzbestimmungen[16] als auch das Lizenzmodell "GeoLizenz"[17] sind vielfältig und unterscheiden sich von den weitverbreiteten Open-Data-Lizenzen wie Creative Commons © erheblich. Die Entwicklung spezieller Lizenzmodelle steht auch im Widerspruch der PSI-Richtline, Standardlizenzen zu verwenden. Die entwickelten Lizenzmodelle unterscheiden beide zwischen kommerziellen und nicht-kommerziellen Zwecken. Das oben beschriebene Beispiel, „offene Daten“ aus den Verwaltungen für einen Datenkonsumenten in der Verwaltung durch einen wirtschaftlich agierenden Dienstleister aufbereiten zu lassen, kann in der Realität nicht praktiziert werden. Eine kombinatorische Auswertung innerstaatlicher oder grenzüberschreitender Verwaltungsdaten bedeutet derzeit, dass in Abhängigkeit des Anwendungskontexts nur ein Bruchteil der verfügbaren Daten zur Generierung eines Informationsproduktes verwendet werden können. Anhand des o. g. Beispiels wird deutlich, dass die Kosten in den Verwaltungen für das Datenmanagement so nicht gesenkt werden können und das Potenzial von Open-Data für die (Geo)Informationswirtschaft sich auch weiterhin nur für theoretische Überlegungen

[14] XRepository, Bundesstelle für Informationstechnik, http://tinyurl.com/d7patwn (31.01.2013)
[15] Gesellschaft für Informatik e.V. (GI) (2010), http://tinyurl.com/cuqx9nb (31.01.2013)
[16] BMI-Datenlizenz Deutschland (2012), http://tinyurl.com/alnwmxv (31.01.2013)
[17] GeoLizenz (2012), http://tinyurl.com/cmhjq74 (31.01.2013)

eignet. Aus Sicht der Autoren müssen Lizenzmodelle für alle offenen Verwaltungsdaten mindestens einem europäischen Kontext folgen, um insbesondere im Umweltkontext Mehrwertdienste durch die (Geo)Informationswirtschaft bereitzustellen. Weiterhin muss geklärt werden, wie Daten mit einer Nutzungsbeschränkung überhaupt in virtuellen Public-Cloud-Umgebungen geschützt gespeichert werden können. Die physische Speicherung von Daten außerhalb Deutschlands erfordert das Akzeptieren der dort vorherrschenden gesetzlichen Rahmenbedingungen[18]. Vor dem Hintergrund der derzeitigen Entwicklung einer „European Union Public Licence" (EUPL) für Open-Source-Software kann die mittelfristige Existenz einer europäischen Konsenslizenz für Open-Data als realistisch eingeschätzt werden.

4.2 Aus Sicht der Wissenschaft

Wie andere Teilgebiete der Informatik verfügt auch die Geoinformatik über ein umfangreiches Portfolio an seriellen Algorithmen, die den steigenden Datenmengen nicht mehr genügen. Anwendungsdisziplinen wie die Physik oder Chemie nutzen seit Jahrzehnten parallele Verfahren für wissenschaftliche Berechnungen, die jedoch nur begrenzt auf die Geoinformatik übertragbar sind. Damit die (Geo)Informationswirtschaft vom Informationsgehalt großer Datenbestände oder kontinuierlicher Datenströme mit Raumbezug profitieren, müssen effiziente Methoden zur parallelen Massendatenverarbeitung entwickelt werden. Auf dieser Grundlage können im Anschluss durch die Geoinformationswirtschaft Mehrwerte für Verwaltung und Gesellschaft geschaffen werden.

Zur Stärkung der Wissensgesellschaft und zur Erzielung nachhaltiger positiver Effekte in der (Geo)Informationswirtschaft sind neue Ausbildungsinhalte wie Datenwissenschaften (Data Science) an Universitäten im Aufbau oder noch zu schaffen. Auch fachübergreifende Forschungskooperationen müssen zunehmend intensiviert werden, da die Schnittmengen der klassischen Ausbildungsinhalte in der Informatik, Umweltwissenschaften etc. sich stetig vergrößern.

4.3 Aus Sicht der Verwaltung

Die Open-Government-Data-Initiativen erfordern eine spezielle IT-Infrastruktur sowie ein strukturiertes und effizientes Datenmanagement, damit auf Basis konsistenter Verwaltungsdaten Mehrwerte in Gesellschaft, Wirtschaft, Wissenschaft und Verwaltung entwickelt werden können. Die Effizienz von Clouddiensten (XaaS) ist im modernen Verwaltungshandeln unbestritten, erfordert aber zusätzliche finanzielle und personelle Ressourcen. Beispielsweise stehen in vielen Verwaltungen dedizierte IT-Infrastrukturabteilungen den einzelnen Abteilungen zur Verfügung. Diese Infrastrukturabteilungen werden in der Regel von mehreren verwaltungsinternen Geschäftsprozessen verwendet. Das Herauslösen und Verlagern eines Geschäftsprozesses in die Cloud bedeutet nicht zwangsläufig, dass die Basisinfrastruktur der bereitstellenden IT-Infrastrukturabteilung gänzlich ersetzt werden kann. Eine Wirtschaftlichkeitsberechnung des beschriebenen Konzepts kann bei größeren Verwaltungen deshalb sehr komplex werden.

Der parallele Aufbau und Betrieb einer Behörden-Cloud wie im Nachbarland Schweiz[19] würde helfen, die Konsolidierung und Restrukturierung von verwaltungsinternen Ge-

[18] Vgl. Foreign Intelligence Surveillance Act Amendment Act (FISAAA 2008)
[19] Cloud Strategie der Schweizer Behörden (2012), http://tinyurl.com/anektgr (31.01.2013)

schäftsprozessen zu systematisieren, erste Synergien zu quantifizieren und Best-Practise-Lösungen abzuleiten. Eine zunehmende Vernetzung der Verwaltung mit der Wirtschaft und Wissenschaft könnte sich eignen, um innovative cloudoptimierte Lösungen zur Optimierung von Verwaltungsprozessen zu schaffen. Damit Lösungen Praxisreife erfahren, muss die domänenspezifische Vernetzung bspw. durch zielgruppenspezifische Konferenzen, Wettbewerbe[20] oder offene Dialoge weiter intensiviert werden.

Literatur

CAO, Y., CHEN, C., GUO, F., JIANG, D., LIN, Y., OOI, B., VO, H., WU, S. & XU, Q. (2011), ES²: A cloud data storage system for supporting both OLTP and OLAP. In: Proceedings of the 27th IEEE conference on Data Engineering (ICDE), 291-302. DOI: 10.1109/ICDE.2011.5767881.

FRAUNHOFER-INSTITUT FOKUS (2012), Studie zur Entwickung einer Open-Data-Strategie für Köln.

KAUR, H., AGRAWAL, P. & DHIMAN, A. (2012), Visualizing clouds on different stages of DHW, An introduction to data warehouses as a service. In: Proceedings of the International Conference on Computing Sciences (ICCS); 356-359; DOI: 10.1109/ICCS. 2012.78.

KYUNG, M.-J, YOM, J.-H. & KIM, S.-Y. (2012), Spatial data warehouse design and spatial OLAP implementation for decision making of geospatial data update. KSCE Journal of Civil Engineering,16, 1023-1031. DOI: 10.1007/s12205-012-1410-2.

MALINOWSKI, E. & ZIMANYI, E. (2011), Advanced Data Warehouse Design, From conventional to spatial and temporal applications. Springer.

MOHAMMAD, H., RAJBABIFARD, A. & WILLAMSON, I. P. (2010), Development of an interoperable tool to facilitate spatial data integration in the context of SDI. IJGIS, 24 (4).

SCHEPERS, T .G. J., IACOB, M. E. & VAN ECK, P. A. T. (2008), A lifecycle approach to SOA governance. In: SAC '08 Proceedings of the 2008 ACM symposium on Applied computing, 1055-1061.

YANG, C., GOODCHILD, M., HUANG, Q., NEBERT, D., RASKIN, R., XU, Y., BAMBACUS, M. & FAY, D. (2011), Spatial Cloud Computing: How geospatial science could use and help to shape cloud computing. International Journal on Digital Earth.

[20] Open-Data-Wettbewerb (2011), http://tinyurl.com/awxr2t7 (31.01.2013)