Lfd.Nr. 4045457

#Big Data in #Austria

Österreichische Potenziale und Best Practice für Big Data

Dr. Martin Köhler, AIT Mobility
Mario Meir-Huber, IDC Central Europe GmbH
Wien, April 2014

Bundesministerium für Verkehr, Innovation und Technologie

Danksagung

Die Studie #Big Data in #Austria wurde im Rahmenprogramm „IKT der Zukunft“ der Österreichischen Forschungsförderungsgesellschaft (FFG) vom Bundesministerium für Verkehr, Innovation und Technologie (BMVIT) beauftragt und gefördert. Wir bedanken uns bei den Fördergebern und bei folgenden Personen und Institutionen für deren wertvollen Beitrag welcher die erfolgreiche Durchführung der Studie ermöglicht hat: Diskussionsleitern des Startworkshops, Teilnehmern des Startworkshops, allen Teilnehmern der Umfragen, allen Interviewpartnern und bei den Teilnehmern der Disseminationsworkshops für wertvolle Kommentare, für wertvollen Input von Partnern aus der Wirtschaft und der Forschung in Bezug auf Informationen zu Anforderungen, Potenzialen und insbesondere zu konkreten Projektumsetzungen. Im Besonderen bedanken wir uns für die Unterstützung durch die Österreichische Computer Gesellschaft (OCG), den IT Cluster Wien, den IT Cluster Oberösterreich, Kollegen der IDC und des AIT, DI (FH) Markus Ray, Dr. Alexander Wöhrer, Prof. Peter Brezany, Dipl. Ing. Mag. Walter Hötzendorfer, Thomas Gregg, Gerhard Ebinger und Prof. Siegfried Benkner.

Projektpartner

IDC Central Europe GmbH
Niederlassung Österreich
Währinger Straße 61
1090 Wien

Austrian Institute of Technology GmbH
Mobility Department, Geschäftsfeld Dynamic Transportation Systems
Giefinggasse 2
1210 Wien

IDC Österreich bringt als international führendes Marktforschungsinstitut langjährige Erfahrung in der Marktanalyse, Aufbereitung und Sammlung von Wissen, Trendanalysen sowie deren Dissemination mit, welche für das Projekt essenziell sind. IDC hat bereits eine Vielzahl von Studien über Big Data erstellt, welche in diese Studie einfließen.

AIT Mobility bringt als innovatives und erfolgreiches Forschungsunternehmen einerseits langjährige Forschungs- und Lehrerfahrung im Bereich Big Data ein und liefert andererseits eine breite thematische Basis für domänenspezifische Analysen (z. B. im Verkehrsbereich). Darüber hinaus hat AIT langjährige Erfahrung in der Durchführung und Konzeption von wissenschaftlichen Studien.

Autoren

Mario Meir-Huber ist Lead Analyst Big Data der IDC in Zentral- und Osteuropa (CEE). In seiner Rolle beschäftigt er sich mit dem Big Data Markt in dieser Region und verantwortet die IDC Strategie zu diesem Thema. Hierfür befasst er sich nicht nur mit innovativen Konzepten und Technologien, sondern teilt sein Wissen mit Marktteilnehmern und Anwendern in diesem Feld. Vor seiner Zeit bei IDC war Mario Meir-Huber bei mehreren Unternehmen in führenden Rollen tätig und hat 2 Bücher über Cloud Computing verfasst. Für die Österreichische Computer Gesellschaft (OCG) ist er der Arbeitskreisleiter für Cloud Computing und Big Data.

Dr. Martin Köhler ist Scientist am Mobility Department der AIT Austrian Institute of Technology GmbH und beschäftigt sich mit Forschungsfragen bezüglich der effizienten Anwendung von Big Data und Cloud Computing Technologien im Mobilitätsbereich. Martin Köhler hat durch seine Mitarbeit an internationalen sowie nationalen Forschungsprojekten langjährige Forschungserfahrung in diesen Themenbereichen und ist Autor von zahlreichen wissenschaftlichen Publikationen. Neben dieser Tätigkeit ist er Lektor für Cloud Computing und Big Data an den Fachhochschulen Wiener Neustadt, St. Pölten und Technikum Wien sowie an der Universität Wien. Seit Herbst 2013 ist er einer der Leiter der OCG Arbeitsgruppe „Cloud Computing und Big Data“.

INHALTSVERZEICHNIS

Zusammenfassung

Die zur Verfügung stehende Datenmenge in den unterschiedlichsten Unternehmen wächst kontinuierlich - Prognosen sprechen von einem durchschnittlichen Wachstum von 33,5% in Umsatzzahlen in den nächsten vier Jahren, wobei der Big Data Markt in Österreich von 22 Millionen Euro im Jahr 2013 auf 73 Millionen Euro im Jahr 2017 anwachsen wird. Dieses rasante Wachstum der verfügbaren Daten führt zu neuen Herausforderungen für ihre gewinnbringende und effiziente Nutzung. Einerseits sind die Daten oft komplex und unstrukturiert, liegen in den unterschiedlichsten Formaten vor und deren Verarbeitung muss zeitsensitiv erfolgen. Andererseits muss das Vertrauen in die Daten sowie in das extrahierte Wissen sichergestellt werden. Diese Problemstellungen werden oft als die vier V zusammengefasst: **Volume, Velocity, Variety -> Value**.

Diese enorme Innovationskraft von Big Data Technologien, von der Datenflut bis hin zu darauf beruhenden semantischen oder kognitiven Systemen, wird in dieser Studie **im Kontext des österreichische Markts analysiert und die entstehenden Potenziale** in unterschiedlichen Sektoren aufgezeigt. Ein wichtiger Ansatz hierbei ist es die Chancen durch die Verwendung von verfügbaren internen und öffentlichen Daten, z. B. Open Government Data, mit einzubeziehen. Dafür wird ein ganzheitlicher Ansatz von der Identifikation der wissenschaftlichen Grundprinzipien, der Entwicklung von neuen Technologien/Methoden/Tools, der Anforderungen, der Umsetzung dieser in innovativen Produkten durch die Wirtschaft bis zur Wissensvermittlung durch Fachhochschulen und Universitäten verfolgt.

Die Studie stellt den **State-of-the-Art in Big Data** dar, definiert den Big Data Stack für eingesetzte Technologien (Utilization, Analytics, Platform und Management Technologien), und gibt einen Überblick über vorhandene Methoden in allen Bereichen des Stacks. Die Aufbereitung des State-of-the-Art inkludiert die umfassende Analyse der österreichischen Marktteilnehmer, der öffentlich zugänglichen Datenquellen sowie der derzeitigen Situation in der tertiären Bildung und Forschung. Weiterführend entwickelt die Studie eine **Markt- und Potenzialanalyse für den österreichischen Markt** welche einen Überblick über die Entwicklung des Big Data Markts gibt, eine detaillierte Evaluierung in Bezug auf Anforderungen, Potenziale und Business Cases nach Branchen beinhaltet und mit Querschnittsthematiken abschließt. Des Weiteren beinhaltet diese Studie einen **Leitfaden für die Umsetzung und Anwendung von Big Data Technologien** um das Ziel der erfolgreichen Abwicklung und Implementierung von Big Data in Organisationen zu erreichen. Hierfür werden Best-Practices-Modelle anhand von spezifischen Big Data Leitprojekten im Detail erörtert und anschließend wird auf dieser Basis ein spezifischer Leitfaden für die Umsetzung von Big Data Projekten in Organisationen präsentiert. In diesem Rahmen werden ein Big Data Reifegradmodell, ein Vorgehensmodell, eine Kompetenzanalyse, unter anderem des Berufsbilds Data Scientist, sowie eine Referenzarchitektur für die effiziente Umsetzung von Big Data vorgestellt. Abschließend stellt diese Studie **Schlussfolgerungen und Empfehlungen** mit dem Ziel der Stärkung der österreichischen Forschungs- und Wirtschaftslandschaft auf Basis der ganzheitlichen Analyse des Bereichs Big Data in Österreich dar.

Die Vorgehensweise bei der Wissensakquise und Durchführung der Studie beruht auf drei Eckpfeilern: Recherche, qualitative Interviews und Umfragen. Einerseits wird eine fundierte **Bestandsaufnahme der aktuellen Markt-, Forschungs- und Bildungssituation sowie der domänenspezifischen Anforderungen** auf Basis

- thematisch passender nationaler und internationaler Studien (z. B. nicht frei verfügbare IDC Studien im Bereich Big Data, Big Data Studie in Deutschland, F&E Dienstleistung österreichische Technologie-Roadmap-Studien),
- bereits durchgeführter Veranstaltungen und vorhandener Technologieplattformen,

- aktuell bewilligter und durchgeführter österreichischer Forschungsprojekte und
- themenspezifischer Wissensvermittlung im tertiären Bildungssektor durchgeführt.

Die Ergebnisse der Recherche wurden durch **branchenspezifische Interviews sowie Diskussionsgruppen im Rahmen von Workshops und durch Umfragen** zu Anforderungen, Umsetzungen und aktueller Expertise ergänzt und validiert.

Executive Summary

The amount of available data in different domains and companies has been increasing steadily. Recent studies expect a yearly revenue growth of 33.5%. The Austrian market volume of EUR 22 million in 2013 is predicted to reach EUR 73 million in 2017. Due to this enormous growth of available data new challenges emerge in efficiently creating value. On the one hand, these data are frequently complex and unstructured in a wide variety of formats and require time-sensitive computation. On the other hand, it is essential to ensure trustworthiness in data and inferred knowledge. These challenges are commonly known as the four V: **Volume, Velocity, Variety -> Value.**

This study analyses the extraordinary, **innovative potential of Big Data technologies for the Austrian market** ranging from managing the data deluge to semantic and cognitive systems. Moreover, the study identifies emerging opportunities arising from the utilization of publicly available data, such as Open Government Data, and company internal data by covering multiple domains. A holistic approach comprising research principles, development of new methods/tools/technologies, requirements, the implementation in innovative products by industry and especially Austrian SMEs, and knowledge transfer at tertiary education establishments served to achieve these objectives.

This study depicts the state-of-the-art of Big Data, defines the Big Data stack for Big Data technologies (Utilization, Analytics, Platform, and Management Technologies) and gives an overview of existing methods in all the areas of the stack. The preparation of the state-of-the-art includes a comprehensive analysis of the Austrian market players, of the publicly accessible data sources as well as the current situation in tertiary education and research. Further, the study develops a market and potential analysis of the Austrian market, which gives an overview of the development of the Big Data market, a detailed evaluation regarding requirements, potentials, and business cases broken down by industries. It closes with cross-cutting issues. Furthermore, this study includes guidelines for the implementation and application of Big Data technologies in order to reach the goal of successfully handling and implementing Big Data in organizations. For this purpose, best practice models will be discussed in detail based on specific Big Data key projects. On this basis, specific guidelines for the realization of Big Data projects in organizations will be presented. Within this scope, a Big Data maturity model, a procedure model, a competence analysis (of the job description of a Data Scientist, among others) as well as reference architecture for an efficient implementation of Big Data will be presented. Finally, this study will present conclusions and recommendations with the goal of strengthening the Austrian research and economic landscape based on a holistic analysis of the area “Big Data” in Austria.

Knowledge acquisition within this study has been based on the following three main building blocks: investigations, interviews, and surveys. The **current market (research, industry, and tertiary education) situation as well as domain-specific and multidisciplinary requirements** will be analyzed based on

- selected national and international studies (e.g. not freely available IDC Big Data studies, Big Data study from Germany, F&E Dienstleistung Österreichische Technologie-Roadmap study),
- recently held events and available technology platforms,
- currently funded and conducted Austrian research projects,
- specialized knowledge transfer at tertiary education establishments.

The achieved results will be complemented and reinforced by means of **domain-specific interviews, discussions in workshops, and surveys**.

1 Einleitung

Die zur Verfügung stehende Datenmenge in den unterschiedlichsten Branchen wächst kontinuierlich. Verschiedenste Daten von den unterschiedlichsten Datenquellen wie Social Media, Smartphones, E-Government Systemen, Video und Audio Streaming, div. Sensornetzwerke, Simulationen und Open Data Initiativen werden immer häufiger gespeichert, um in weiterer Folge analysiert zu werden. Beispielsweise hat der globale Datenverkehr im Jahr 2013 rund 18 Exabyte betragen, was circa der 18fachen Größe des gesamten Internets im Jahr 2000 entspricht (Cisco, 2014). Zusätzlich lässt der Trend zu Systemen auf Basis von Internet of Things (Atzori, Iera, & Morabito, 2010) oder Industrie 4.0 (Sendler, 2013) die Datenmenge weiter steigen. Dies lässt sich in zahlreichen Bereichen wie Verkehr, Industrie, Energie, Gesundheit, Raumfahrt, Bildung, des öffentlichen Bereichs und auch im Handel beobachten. Dieser Wechsel hin zu der Verfügbarkeit von großen, oft schnelllebigen und unterschiedlichsten Daten erfordert ein Umdenken in der Speicherung, Verarbeitung und Analyse der Daten für ihre gewinnbringende und effiziente Nutzung. Diese neuen Herausforderungen werden oft als die vier V bezeichnet, welche die Größe der Daten, deren Komplexität und Struktur sowie deren zeitsensitive Verarbeitung beschreiben: **Volume, Velocity, Variety -> Value**.
Nichtsdestotrotz birgt die zur Verfügung stehende Datenmenge enormes Potenzial in der Durchführung von Forschung sowie auch für die Umsetzung und Erschließung neuer und innovativer Geschäftsfelder für Unternehmen. In der Forschung ist seit einigen Jahren ein Wechsel zu einem neuen vierten **datengetriebenen Paradigma der Wissenschaft** zu erkennen (Hey, Tansley, & Tolle, 2009), (Critchlow & Kleese Van Dam, 2013). Datenintensive Forschung beschreibt einen neuen Forschungsansatz neben experimenteller, theoretischer Forschung und Computersimulationen von natürlichen Phänomenen (Kell, 2009) auf Basis der verfügbaren großen Datenmengen, deren Verschneidung und neuartiger Interpretation und Analyse. Um diese neuen Ansätze in der Forschung als auch in der Wirtschaft einsetzbar zu machen und die daraus resultierenden Potenziale zu schöpfen, entstehen neue innovative Technologien und Methoden. Durch deren Umsetzung und Anwendung in unterschiedlichen Branchen können innovative Geschäftsfelder erschlossen sowie bestehende optimiert werden. Aus diesem Umfeld entstanden in den letzten Jahren neue Forschungsfelder sowie ein neues Marktumfeld. Dieser Big Data Markt ist in den letzten Jahren international stark gewachsen und aktuelle Prognosen sprechen für Österreich von einem durchschnittlichen Wachstum von 33,5% in Umsatzzahlen in den nächsten vier Jahren, wobei der Big Data Markt in Österreich von 22 Millionen Euro in 2013 auf 73 Millionen Euro in 2017 anwachsen wird (siehe Kapitel 3).

Diese enorme Innovationskraft von Big Data Technologien, von der Datenflut bis hin zu darauf beruhenden semantischen oder kognitiven Systemen, werden in dieser Studie **im Kontext des österreichischen Markts analysiert und die entstehenden Potenziale** in unterschiedlichen Sektoren aufgezeigt. Dafür wird ein ganzheitlicher Ansatz von der Identifikation der wissenschaftlichen Grundprinzipien, der Entwicklung von neuen Technologien/Methoden/Tools, der Anforderungen, der Umsetzung dieser in innovativen Produkten **durch die Wirtschaft bis zur Wissensvermittlung durch Fachhochschulen und Universitäten** verfolgt. Die Studie stellt zuerst den State-of-the-Art in Big Data dar, definiert den Big Data Stack für Big Data Technologien, und gibt einen Überblick über vorhandene Methoden in allen Bereichen des Stacks. Die Aufbereitung des State-of-the-Art inkludiert die umfassende Analyse der österreichischen Marktteilnehmer, der öffentlich zugänglichen Datenquellen sowie der derzeitigen Situation in der tertiären Bildung und Forschung. Weiterführend entwickelt die Studie eine Markt- und Potenzialanalyse für den österreichischen Markt, welche einen Überblick über die Entwicklung des Big Data Markts gibt, eine

detaillierte Evaluierung in Bezug auf Anforderungen, Potenziale und Business Cases nach Branchen beinhaltet und mit Querschnittsthematiken abschließt. Des Weiteren werden anschließend identifizierte Leitprojekte im Detail analysiert und auf dieser Basis ein Leitfaden für die Umsetzung von Big Data Projekten dargestellt. Die Studie schließt mit Schlussfolgerungen und gezielten Empfehlungen für die Stärkung des Wirtschafts- und Forschungsstandorts Österreich in Bezug auf Big Data.

1.1 Ziele des Projektes

Ziel der Studie „#BigData in #Austria“ ist das Aufzeigen von konkreten Handlungsoptionen für Wirtschaft, Politik, Forschung und Bildung anhand von identifizierten Leitdomänen und innovativen Business Cases für den österreichischen Markt. Die Studie bietet einen ganzheitlichen Ansatz, um den Wissens- und Technologietransfer von der Forschungslandschaft bis zur Wirtschaft abzubilden und zu analysieren. Ein zusätzliches Augenmerk wird dabei auf die Einbeziehung der tertiären Bildungssituation gelegt.

Die Studie definiert hierfür den State-of-the-Art im Bereich Big Data, von der Datenflut, deren Analyse bis hin zu darauf beruhenden semantischen und kognitiven Systemen, anhand von wissenschaftlicher Recherchearbeit und analysiert bestehende Verfahren und Technologien. Zusätzlich werden vorhandene öffentlich zugängliche Datenquellen am österreichischen Markt identifiziert und Potenziale des vorhandenen Commitments zu Open Data der österreichischen Politik analysiert.

Die Studie führt eine umfassende Analyse der aktuellen Situation in Bezug auf Big Data in der österreichischen Markt-, Forschungs- und tertiären Bildungslandschaft durch, dargestellt in Abbildung 1. Diese beinhaltet die Untersuchung von gezielten Branchen und von relevanten Business Cases auf Basis von mehreren vorhandenen Studien und Marktforschungsergebnissen der IDC, internationalen Studien in diesem Bereich, nationalen Vorstudien, Recherchearbeit und Interviews mit EntscheidungsträgerInnen entsprechender Branchen. Zusätzlich wird speziell auf die Rolle Österreichs in dem Bereich Open Government Data eingegangen. Basierend auf der eingehenden Analyse der Branche Mobilität sowie der erhobenen Daten werden domänenübergreifende Anforderungen und Potenziale identifiziert und mit Hilfe von Umfragen mit Stakeholdern in den Branchen validiert.

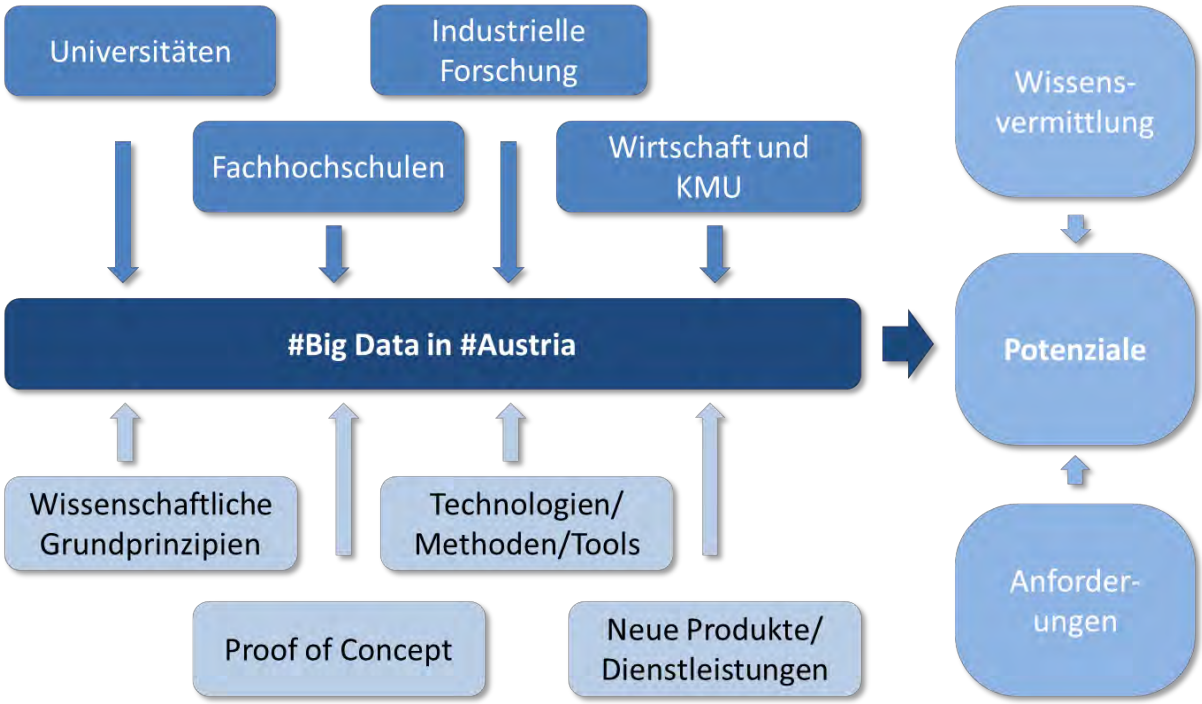


Abbildung 1: Ganzheitliche Sicht auf Big Data in Österreich

Ein wichtiges Ergebnis neben der Evaluierung der technischen Machbarkeit, ist auch die Betrachtung der wirtschaftlichen und politischen Umsetzbarkeit. Hierfür wird das extensive Entscheidernetzwerk der IDC eingebunden, um langfristige Trends und den Bedarf im Bereich Big Data am österreichischen Markt zu identifizieren und zu analysieren. Diese Studie bietet eine umfangreiche Hilfestellung für MarktteilnehmerInnen und potenziell neue Unternehmen, die in eben diesem Bereich in Österreich tätig sind/sein wollen. Wesentlich hierfür ist, dass die Studie einen umfangreichen Handlungsvorschlag für Unternehmen und Organisationen in Österreich im Rahmen eines Leitfadens sowie Empfehlungen und Schlussfolgerungen darstellt. Dies wird durch das bereits sehr umfangreiche Wissen der IDC im Go-to-Market Umfeld und der wissenschaftlichen Expertise des AIT erreicht. Zu diesem Zweck wird auch das in der IT Branche sehr angesehene IDC Reifegradmodel[1]: Big Data and Analytics an die Erfordernisse der Studie angepasst, so dass identifiziert werden kann auf welcher Ebene sich die Technologieentwicklung in einer bestimmten Organisation befindet.

Ein weiteres Ziel des Projekts ist das Aufzeigen der technischen Machbarkeit von Big Data Projekten am Wirtschafts- und Forschungsstandort Österreich. Hierfür werden Big Data Projekte auf Basis von Recherchearbeit und Interviews identifiziert und im Detail auf Anforderungen, Potenziale und Technologien analysiert. Diese Analyse von Projekten resultiert in einem prototypischen Leitfaden für die Abwicklung von Big Data Projekten welches ein Reifegradmodell, ein Vorgehensmodell sowie Empfehlungen zu Datenschutz, Sicherheit und eine Big Data Referenzarchitektur bietet.

Die Ergebnisse beinhalten eine aktuelle und ganzheitliche Sicht auf den Bereich Big Data (siehe Abbildung 1) welche einerseits eine Definition und Abgrenzung des Bereichs und eine Erhebung der vorhandenen technischen Verfahren sowie eine Übersicht und Analyse der öffentlich verfügbaren

[1] Das IDC Reifegradmodell wird für verschiedene Technologien eingesetzt und bietet IT Entscheidungsträgern Informationen über jeweilige Technologien und deren Implementierungen.

Datenquellen bietet, andererseits zusätzlich einen Überblick des Wissens- und Technologietransfers von Forschungseinrichtungen, dem tertiären Bildungssektor bis hin zu lokalen KMUs beinhaltet.

Des Weiteren ist eine detaillierte Markt- und Potenzialanalyse dargestellt, welche die Anforderungen und Potenziale verschiedener Branchen und relevante Business Cases beinhaltet. Zusätzlich schafft diese Analyse die Basis, um konkrete Handlungsoptionen für Wirtschaft, Politik und Forschung für die zukünftige positive Entwicklung und Stimulierung des österreichischen Markts aufzuzeigen.

1.2 Konkrete Fragestellungen

In dieser Studie werden die konkreten Fragestellungen der Ausschreibung im Detail bearbeitet und diskutiert. Nachfolgend sind diese Fragestellungen gelistet und es wird konkret beschrieben, in welchen Abschnitten diese Fragestellungen diskutiert werden.

Was sind domänenspezifische Anforderungen für den erfolgreichen Einsatz von Big Data Verfahren?

Die Studie identifiziert Herausforderungen für den erfolgreichen Einsatz von Big Data und daraus resultierende Potenziale und Business Cases in verschiedenen Branchen. Hierfür werden in Kapitel 3.2.4 siebzehn Branchen im Detail analysiert. Zusätzlich diskutiert die Studie Anforderungen und Potenziale anhand von spezifischen Big Data Projekten, welche in Österreich beziehungsweise mit führender Österreichischer Beteiligung durchgeführt werden. In weiterer Folge werden diese domänenspezifischen Anforderungen und Potenziale in domänenübergreifende Anforderungen anhand von wesentlichen Querschnittsthemen klassifiziert. Diese spiegeln sich auch im erstellten Leitfaden für die erfolgreiche Umsetzung von Big Data Projekten (siehe Kapitel 4.2) wider.

Welche Verfahren und Werkzeuge stehen dafür zur Verfügung?

Die Studie identifiziert und diskutiert den aktuellen State-of-the-Art von Big Data und erarbeitet einen detaillierten Überblick über die vorhandenen Werkzeuge sowie Verfahren in Kapitel 2.2.

Auf welche offenen Daten kann zugegriffen und dadurch Mehrwert erzeugt werden?

Die Studie diskutiert in Kapitel 2.5 die aktuell verfügbaren öffentlich zugänglichen Daten am Österreichischen Markt. Hierbei wird auch ein Fokus auf das Österreichische Commitment zu Open Government Data gelegt und es werden derzeitige Initiativen, rechtliche Implikationen und zukünftige Entwicklungen diskutiert.

Welche Schritte sind durchzuführen, um ein Big Data Projekt aufzusetzen?

Die Studie analysiert mehrere Big Data Leitprojekte aus den in dieser Studie identifizierten Leitdomänen (Verkehr, Gesundheits- und Sozialwesen, Handel) und stellt deren Anforderungen gegenüber. Auf der Basis dieser Analyse (Kapitel 4.1), dem Überblick über technische Verfahren (Kapitel 2.2), sowie der Analyse der österreichischen Marktsituation (Kapitel 2.3 sowie Kapitel 3) entsteht ein prototypischer Leitfaden für die Abwicklung von Big Data Projekten welcher im Detail in Kapitel 4.2 ausgearbeitet ist.

Welche Einschränkungen sind beim Zugriff auf Daten zu beachten?

Die Studie beleuchtet die aktuell öffentlich verfügbaren Datenquellen und gibt einen Überblick über die vorherrschenden Zugriffmöglichkeiten auf und Lizenzen von diesen Daten (Kapitel 2.5). Zusätzlich hat die Studie Einschränkungen der unternehmensinternen Datennutzung in den Umfragen beleuchtet und diskutiert sicherheitstechnische und rechtliche Aspekte in Kapitel 4.2.4.

Welche Anbieter von Werkzeugen und Know-how bestehen?

Die Studie diskutiert die aktuell am österreichischen Markt verfügbaren Werkzeuge und deren AnbieterInnen sowie welche MarktteilnehmerInnen aus der Wirtschaft, Forschung und dem tertiären Bildungsbereich Expertise im Bereich Big Data bereitstellen in Kapitel 2.3 und Kapitel 2.4.

1.3 Methodik und Vorgehen

Die Vorgehensweise bei der Wissensakquise und Durchführung der Studie beruht auf drei Eckpfeilern: Recherche, qualitative Interviews mit Stakeholdern sowie Umfragen. Die fundierte Bestandsaufnahme der verfügbaren Angebote sowie der domänenspezifischen Anforderungen basiert auf thematisch passenden nationalen und internationalen Studien, bereits durchgeführten Veranstaltungen und vorhandenen Technologieplattformen sowie von IDC durchgeführten Marktanalysen. Des Weiteren werden Erkenntnisse aus internen (AIT) sowie externen, bewilligten und durchgeführten österreichischen Forschungsprojekten herangezogen.

Im Detail wurden folgende vorhandene und Big Data Studien in die Ergebnisse miteinbezogen. Hierbei handelt es sich teilweise um frei zugängliche Studien sowie um Studien welche von IDC mit eingebracht wurden:

- Österreich:
  - Berger, et. al., Conquering Data in Austria, Technologie Roadmap für das Programm IKT der Zukunft: Daten durchdringen – Intelligente Systeme, 2014
  - IDC Software Report, 2013
  - Austria IT Services Market 2013–2017 Forecast and 2012 Analysis. August 2013
  - IDC Big Data Survey Austria, 2012 (dediziert für den österreichischen Markt erstellt)
  - IDC Big Data Survey Austria, 2013 (dediziert für den österreichischen Markt erstellt)
- Deutschland:
  - BITKOM, Leitfaden: Management von Big-Data-Projekten, 2013
  - BITKOM, Leitfaden Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte, 2013
  - BITKOM, Leitfaden Big-Data-Technologien – Wissen für Entscheider, 2014
  - Markl, et. al., Innovationspotenzialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen (Big Data Management), 2014
  - Fraunhofer, Big Data – Vorsprung durch Wissen: Innovationspotenzialanalyse, 2012
- International:
  - NESSI White Paper, Big Data: A New World of Opportunities, 2012
  - Big Data Public Private Forum, Project Deliverables, 2013
  - TechAmerica Foundation, Demystifying Big Data, 2012
  - Interxion, Big Data – Jenseits des Hypes, 2013
  - EMC, Harnessing the Growth Potential of Big Data, 2012
  - IDC Worldwide Big Data Technology and Services 2013–2017 Forecast, Dec 2013
  - IDC Worldwide Storage in Big Data 2013–2017 Forecast, May 2013
  - IDC's Worldwide Storage and Big Data Taxonomy, 2013, Feb 2013
  - Big Opportunities and Big Challenges: Recommendations for Succeeding in the Multibillion-Dollar Big Data Market, Nov 2012
  - IDC Predictions 2013: Big Data – Battle for Dominance in the Intelligent Economy, Jan 2013
  - IDC Maturity Model: Big Data and Analytics — A Guide to Unlocking Information Assets, Mar 2013
  - IDC Maturity Model Benchmark: Big Data and Analytics in North America, Dec 2013

  - Big Data in Telecom in Europe, 2013: Enhancing Analytical Maturity with New Technologies, Dec 2013

Des Weiteren wurden Ergebnisse der folgenden themenrelevanten Studien für die Ergebnisse dieser Studie evaluiert:

- Österreich:
  - EuroCloud, Leitfaden: Cloud Computing: Recht, Datenschutz & Compliance, 2011
  - IT Cluster Wien, Software as a Service – Verträge richtig abschließen 2., erweiterte Auflage, 2012
  - WU Wien, Die wirtschaftspolitische Dimension von Open Government Data in Österreich, 2012
  - Eibl, et. al., Rahmenbedingungen für Open Government Data Plattformen, 2012

Im Rahmen eines Startworkshops zu „#BigData in #Austria“ wurde am Donnerstag den 12. Dezember 2013 von 9:30 bis 14:00 im MEDIA TOWER, Taborstraße 1-3, 1020 Wien über die Studie und deren Ziele informiert und im Anschluss wurden erste Erhebungen im World Café Stil durchgeführt. Es wurden die Themen Big Business, Big Challenges, Big Science und Big Datenschutz diskutiert.

Auf Basis der aus der breiten verfügbaren Datenbasis gewonnenen Informationen, der bereits von IDC durchgeführten Befragungen (siehe beschriebene Datenbasis) und der Ergebnisse des Startworkshops wurde ein Interviewleitfaden für die Ergänzung der Informationen bezüglich der österreichischen Marktsituation (ganzheitlich: Forschung, Industrie, Bildung, Politik – siehe Abbildung 1) erstellt und ein valider Pool an EntscheidungsträgerInnen für Interviews und Umfragen ausgewählt und befragt. Ein Überblick über die durchgeführte Methodik ist in Abbildung 2 dargestellt.

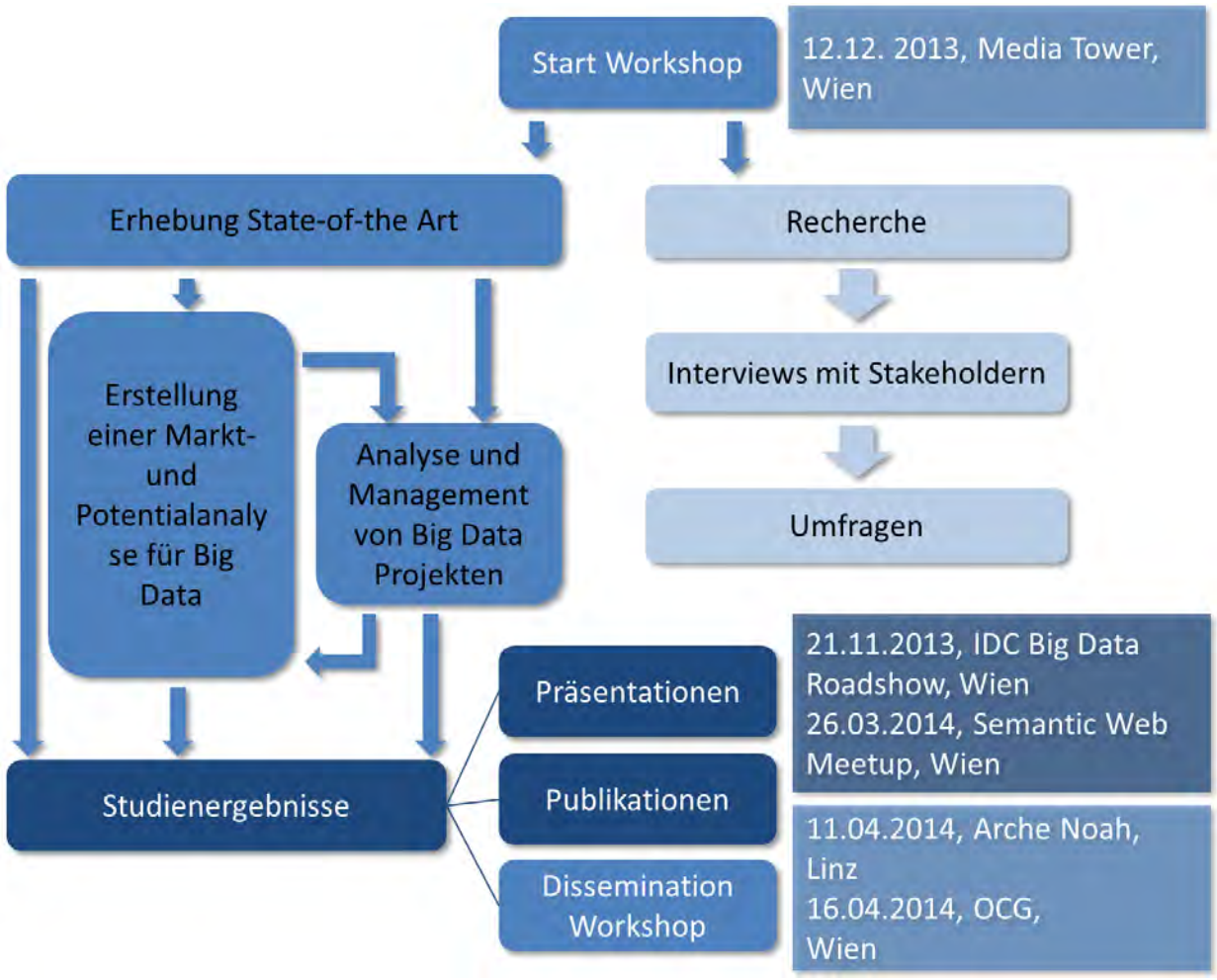


Abbildung 2: Methodik der Studie #Big Data in #Austria

Während der Erstellung der Markt- und Potenzialanalyse für Österreich wurden auf Basis des State-of-the-Art im Bereich Big Data, der Ergebnisse des Startworkshops sowie der durchgeführten Stakeholder-Befragungen die Branchen, Business Cases und Projekte identifiziert und es wurde eine Analyse von Querschnittsthematiken erstellt. Für die Analyse und das Management von Big Data Projekten wurde auf Rechercheergebnisse sowie auf Ergebnisse der Marktanalyse zurückgegriffen, um auf Basis abgewickelter sowie laufender Projekte einen prototypischen Leitfaden für Big Data Projekte zu erstellen.

Die Ergebnisse der Studie wurden in Präsentationen und zwei Dissemination Workshops, in enger Zusammenarbeit mit dem IT Cluster Linz und der OCG, Stakeholdern aus der Wirtschaft und Forschung im Detail präsentiert und diskutiert. Die Ergebnisse der Diskussionen wurden in der Folge in die Studienresultate eingearbeitet. Eine detailliertere Beschreibung der Workshops kann dem Anhang entnommen werden.

1.4 Struktur des Berichts

Der Endbericht der Studie „#BigData in #Austria“ gliedert sich in drei Hauptabschnitte. Dieser Bericht beginnt mit einer detaillierten Analyse des State-of-the-Art von Big Data in Kapitel 2. Zuallererst wird der Bereich Big Data im Detail definiert und dessen Dimensionen sowie technologischen Implikationen werden aufgezeigt. Anschließend folgt eine Aufschlüsselung der für diesen Themenbereich relevanten und vorhandenen Verfahren und Technologien. Weiters wird der

Österreichische Markt in Bezug auf Unternehmen, Forschung und tertiäre Bildung diskutiert und abschließend werden der aktuelle Status in Bezug auf Open Government Data abgebildet.

Der zweite Teil des Endberichts (siehe Kapitel 3) widmet sich einer detaillierten Markt und Potenzialanalyse. Nach einer Darstellung der internationalen Marktsituation wird die europäische Situation diskutiert und anschließend auf österreichische Spezifika eingegangen. Darauf folgt eine detaillierte Darstellung von Herausforderungen, Potenzialen und Business Cases in verschiedenen Branchen des österreichischen Markts. Nach diesen domänenspezifischen Anforderungen werden Querschnittsthemen unabhängig von der jeweiligen Branche analysiert.

Der dritte Teil der Studie (siehe Kapitel 4) befasst sich mit der Darstellung von Best Practices für Big Data für Österreichische Unternehmen und Organisationen. Hierbei werden zuallererst Vorzeigeprojekte in Bezug auf Big Data Charakteristiken, Big Data Technologien, Anforderungen und Potenziale analysiert. Darauf folgend wird ein Leitfaden für die effiziente Umsetzung von Big Data Projekten vorgestellt. Der Leitfaden beinhaltet ein Reifegradmodell für Big Data anhand dessen Organisationen ihren aktuellen Status in Bezug auf Big Data einschätzen können. Anschließend wird anhand eines Vorgehensmodells in Bezug auf die Abwicklung von Big Data Projekten auf diesbezügliche Spezifika eingegangen. Der Leitfaden wird mit einer detaillierten Kompetenzanalyse in Bezug auf Big Data (Stichwort „Data Scientists“), einer Diskussion der wichtigen Thematik Datenschutz und Sicherheit sowie mit einer sorgfältigen Ausführung einer Referenzarchitektur für Big Data abgeschlossen.

Die Studie schließt mit einer ausführlichen Diskussion mit Schlussfolgerungen und Empfehlungen zur Stärkung des Standorts Österreich in Bezug auf Big Data. Ab Anhang A werden Inhalte der abgehaltenen Workshops näher erläutert.

2 State-of-the-Art im Bereich Big Data

In diesem Kapitel wird der aktuelle Status des Bereichs Big Data in Österreich umfassend aufbereitet und eine ganzheitliche Analyse des Bereichs durchgeführt. Nach der grundlegenden Definition dieses Bereichs, wird ein Big Data Technologie Stack definiert und dieser von anderen Bereichen abgegrenzt. Anschließend werden Österreichische MarktteilnehmerInnen identifiziert und deren Kompetenz im Bereich Big Data analysiert. Darüber hinaus wird das Angebot in den Bereichen tertiäre Bildung und Forschung dargestellt und abschließend werden öffentlich verfügbare Datenquellen und deren Verwendbarkeit für Forschung und Wirtschaft dargelegt.

2.1 Definition und Abgrenzung des Bereichs „Big Data“

Die zur Verfügung stehende Datenmenge aus den unterschiedlichsten Datenquellen steigt in den letzten Jahren stetig. Daten von verschiedensten Sensoren, Social Media Technologien, Smartphones, E-Government Systemen, Video und Audio Streaming Lösungen, Simulationen und Open Data Initiativen müssen für die Verarbeitung oftmals gespeichert werden. Zusätzlich lässt der Trend zu Systemen auf Basis von Internet of Things oder Industrie 4.0 die Datenmenge weiter steigen. Diese Entwicklung kann in vielen Bereichen wie Verkehr, Industrie, Energie, Gesundheit, Raumfahrt, Bildung, des öffentlichen Bereichs und auch im Handel beobachtet werden. Der Wechsel hin zur Verfügbarkeit von großen, oft schnelllebigen und unterschiedlichsten Daten erfordert ein Umdenken in der Speicherung, Verarbeitung, Analyse und Wissensgenerierung aus Daten. Unter dem Begriff Big Data wird generell die Extraktion von neuem Wissen für die Unterstützung von Entscheidungen für unterschiedlichste Fragestellungen auf Basis des steigenden heterogenen Datenvolumens verstanden. Dieses neue Wissen bildet den Mehrwert für Organisationen auf Basis des neuen Rohstoffes – den Daten.

2.1.1 Big Data Definitionen

Gemeinhin wird der Begriff Big Data unterschiedlich definiert und zusammengefasst. Die Networked European Software and Services Initiative (NESSI), eine europäische Technologieplattform für Industrie und Wissenschaft definiert Big Data als einen Begriff, welcher die Verwendung von Technologien für die Erfassung, die Verarbeitung, die Analyse und die Visualisierung von großen Datenmengen unter Rücksichtnahme auf die Zugriffszeit umfasst *(NESSI, 2012)*:

> *““Big Data” is a term encompassing the use of techniques to capture, process, analyse and visualize potentially large datasets in a reasonable timeframe not accessible to standard IT technologies. By extension, the platform, tools and software used for this purpose are collectively called “Big Data technologies”.*

Der deutsche IT-Branchenverband Bitkom definiert den Begriff Big Data in BITKOM, 2012 so:

> *„Big Data stellt Konzepte, Technologien und Methoden zur Verfügung, um die geradezu exponentiell steigenden Volumina vielfältiger Informationen noch besser als fundierte und zeitnahe Entscheidungsgrundlage verwenden zu können und so Innovations- und Wettbewerbsfähigkeit von Unternehmen weiter zu steigern.“*

Im Rahmen des Bitkom Big Data Summits 2014 hat Prof. Dr. Stefan Wrobel Big Data wie folgt definiert *(Wrobel, 2014)*:

> *„Big Data bezeichnet allgemein*
>
> - *Den Trend zur Verfügbarkeit immer detaillierterer und zeitnäherer Daten*
> - *Den Wechsel von einer modellgetriebenen zu einer daten-und modellgetriebenen Herangehensweise*
> - *Die wirtschaftlichen Potenziale, die sich aus der Analyse großer Datenbestände bei Integration in Unternehmensprozesse ergeben*
>
> *Big Data unterstreicht aktuell technisch folgende Aspekte*
>
> - *Volume, Variety, Velocity*
> - *In-memory computing, Hadoop etc.*
> - *Real-time analysis und Skaleneffekte*
>
> *Big Data muss gesellschaftliche Aspekte zentral mit berücksichtigen.“*

Daraus wird abgeleitet, dass Big Data somit viele unterschiedliche Aspekte, von der steigenden Menge an heterogenen Daten, der Verarbeitung und Analyse dieser in Echtzeit, bis hin zu der Wissens- beziehungsweise Mehrwertgenerierung aus den Daten, umfasst. Gemeinhin werden diese Charakteristiken auch unter den vier V zusammengefasst welche öfters durch *Veracity* (Vertrauen in die Daten) und *Visualization* (Visualisierung der Daten) erweitert werden.

- **Volume**. Bezeichnet den enormen Anstieg der vorhandenen Datenmenge in den letzten Jahren. Oftmals werden bis zu Petabyte an Daten produziert. Die Herausforderungen bestehen in der Verwaltung dieser Daten und in der effizienten Ausführung von Analysen auf dem Datenbestand.
- **Variety**. Verschiedene Datenquellen liegen in unterschiedlichen Formaten vor und sind oft komplex und unstrukturiert. Daten werden oft in unterschiedlichen Formaten zurückgegeben. Dies umfasst arbiträre Datenformate, strukturierte Daten (zum Beispiel relationale Daten) bis hin zu komplett unstrukturiertem Text. Die Herausforderung für Anwendungen ist die flexible Integration von Daten in unterschiedlichsten Formaten.
- **Velocity**. Daten müssen oft direkt verarbeitet werden und Ergebnisse sollen zeitnah zur Verfügung stehen. Immer mehr Sensordaten werden abgefragt und produziert, welche auch für vielfältige Anwendungszwecke in Echtzeit analysiert werden müssen. Das stellt eine große Herausforderung für Anwendungen dar.
- **Value**. Value beinhaltet das Ziel der gewinnbringenden Nutzung der Daten. Daten sollen schlussendlich auch einen gewissen Mehrwert für das Unternehmen oder Organisation bringen.

Somit ergibt sich als klares Ziel des Bereichs Big Data Mehrwert auf Basis des neuen Rohstoffes, den Daten, zu generieren und damit entstehende Potenziale innerhalb von Organisationen auszuschöpfen. Um dieses Ziel zu erreichen werden innovative neue Methoden benötigt, welche unter dem Begriff Big Data Technologien zusammengefasst werden. Big Data Technologien umfassen alle Hardware und Software Technologien für das Management, die Analyse von und die Mehrwertgenerierung aus großen heterogenen Datenmengen.

Immer wieder entstehen Diskussionen ab welcher Datenmenge es sich um Big Data Problematiken handelt. In dieser Studie wird Big Data in dieser Hinsicht wie folgt definiert: Von Big Data wird gesprochen sobald herkömmliche Lösungen für die Verarbeitung und Mehrwertgenerierung auf Basis der Daten nicht mehr ausreichen und diese an ihre Grenzen stoßen. In diesem Falle müssen aufgrund der vorhandenen Daten (Volume, Velocitiy, Variety) neue Big Data Technologien eingesetzt werden. Dieses Paradigma kann einfach wie folgt zusammengefasst werden:

„Big Data: Wenn die Daten selbst zu dem Problem werden, um einen Mehrwert daraus abzuleiten zu können."

2.1.2 Big Data Dimensionen

Gerade im Kontext der aktuellen Entwicklungen muss Big Data von mehreren Dimensionen aus betrachtet werden. Neben der technologischen Dimension (welche Technologien werden für den erfolgreichen Einsatz von Big Data benötigt), welche in dieser Studie im Detail behandelt wird, müssen auch juristische, soziale, wirtschaftliche und anwendungsspezifische Aspekte miteinbezogen werden. Die juristischen Aspekte umfassen Datenschutz, Sicherheit sowie vertragliche Aspekte. Gerade Datenschutz spielt eine wesentliche Rolle in der erfolgreichen Umsetzung von Big Data. Dem Umgang mit personenbezogenen Daten muss gesonderte Beachtung geschenkt werden und hier gilt es, rechtliche Rahmenbedingungen und technologische Möglichkeiten für dessen Umsetzung zu berücksichtigen. Die Verwendung von Big Data kann auch wichtige soziale Implikationen hervorrufen. Auswirkungen auf die Gesellschaft sowie auf das Benutzerverhalten gegenüber einzelnen Unternehmen sowie im aktuellen digitalen Umfeld sind wichtige Aspekte welche in der Umsetzung beachtet werden müssen. Bei der Umsetzung von Big Data in Unternehmen sind die Ziele die Ausnutzung der Potenziale und die Generierung von Nutzen. Hierbei werden spezifische Business Modelle sowie innovative Preisgestaltungsmechanismen benötigt. Abhängig von dem jeweiligen Sektor beziehungsweise der Anwendung gilt es unterschiedliche applikationsspezifische Anforderungen und Herausforderungen in Bezug auf Big Data zu meistern. Als Beispiele werden hier die Echtzeitverarbeitung von anonymisierten Sensordaten in der Verkehrsdomäne sowie die Verwendung von Big Data im aufkommenden Bereich Industrie 4.0 genannt. Ein Überblick über die Dimensionen von Big Data wird in Abbildung 3 dargestellt.

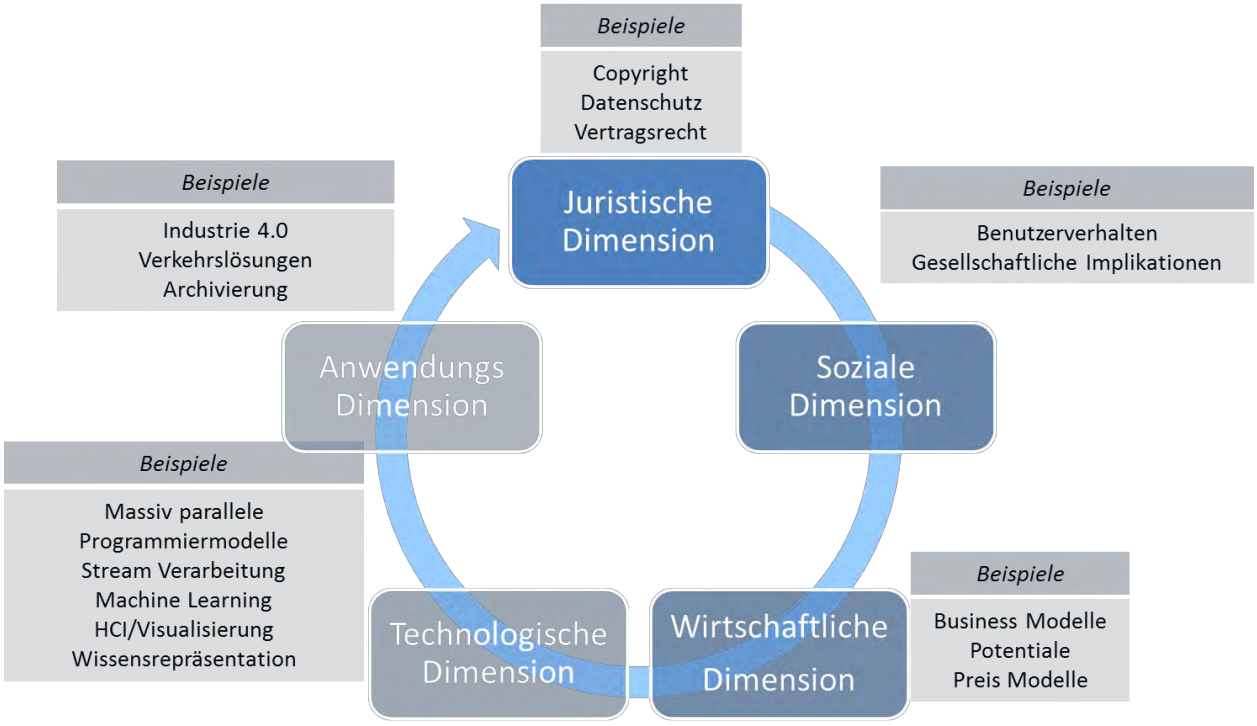


Abbildung 3: Dimensionen von Big Data (Markl, Big Data Analytics – Technologien, Anwendungen, Chancen und Herausforderungen, 2012)

In dieser Studie wird ein Fokus auf die technologische Dimension gesetzt und diese über den nachfolgenden Big Data Stack definiert (detaillierte Analyse siehe Kapitel 2.2). Des Weiteren werden die Anwendungs- und wirtschaftliche Dimensionen in Abschnitt 3.2.4 sowie die rechtliche und soziale Dimensionen in Kapitel 4 behandelt.

2.1.3 Der Big Data Stack

Eine ganzheitliche Sicht - unter Berücksichtigung aller V's - auf Big Data ergibt ein breites Spektrum an technologischen Lösungen von der Bereitstellung von IT-Infrastruktur bis hin zur Visualisierung der Daten. Diese Studie definiert und kategorisiert Big Data-relevante Technologien im Folgenden anhand des Big Data Stacks. Dieser wird, wie in Abbildung 4 ersichtlich, in vier Ebenen eingeteilt: Management, Platform, Analytics und Utilization, welche im Folgenden näher erläutert werden.

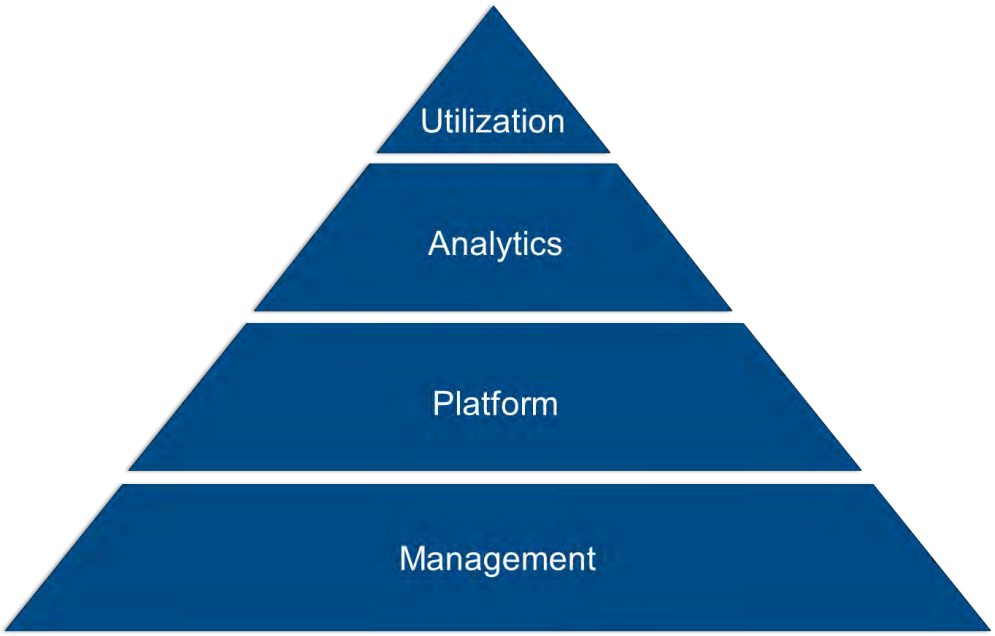


Abbildung 4: Big Data Stack

Utilization

Big Data Technologien und Verfahren werden von Unternehmen und Forschungseinrichtungen unabhängig von dem Geschäftsfeld mit einem bestimmten Ziel angedacht: Sie sollen Mehrwert generieren und dadurch die aktuelle Marktsituation der Organisation stärken. Dies wird häufig unter dem vierten V, Value, zusammengefasst. Die Ebene Utilization beschäftigt sich mit allen Technologien, welche dieses Ziel der Nutzbarmachung der Daten verfolgen. Als beide werden hierfür Wissensmanagementsysteme oder Visualisierungstechnologien genannt.

Analytics

Der Bereich Analytics beschäftigt sich mit der Informationsgewinnung aus großen Datenmengen auf Basis von analytischen Ansätzen. Das Ziel dieser Ebene ist die Erkennung von neuen Modellen auf Basis der zur Verfügung stehenden Daten, das Wiederfinden von vorhandenen Mustern oder auch die Entdeckung von neuen Mustern. Wichtige Technologien inkludieren Machine Learning sowie Methoden aus der Mathematik und Statistik.

Platform

Die Platform-Ebene beschäftigt sich mit der effizienten Ausführung von Datenanalyseverfahren auf großen Datenmengen. Um analytische Verfahren effizient ausführen zu können, werden skalierende und effiziente Softwareplattformen für deren Ausführung benötigt. Die Hauptaufgaben hierbei sind die Bereitstellung von parallelisierten und skalierenden Ausführungssystemen sowie die Echtzeitanbindung von Sensordaten.

Management

Management umfasst Technologien, welche die effiziente Verwaltung von großen Datenmengen ermöglichen. Dies umfasst die Speicherung und Verwaltung der Datenmengen für effizienten Datenzugriff sowie die Bereitstellung und das Management der darunter liegenden Infrastruktur.

2.2 Klassifikation der vorhandenen Verfahren und Technologien

In diesem Kapitel werden vorhandene Verfahren und Technologien auf Basis der vier Technologieebenen klassifiziert und in Bezug auf Big Data detailliert betrachtet. Diese Studie erhebt in diesem Zusammenhang keinen Anspruch auf Vollständigkeit, versucht aber den aktuellen Stand der Technik und Wissenschaft in einer geeigneten Darstellung widerzuspiegeln. Abbildung 5 gibt einen Überblick über die vier Ebenen und die darin inkludierten Technologien welche in der weiteren Folge näher analysiert werden.

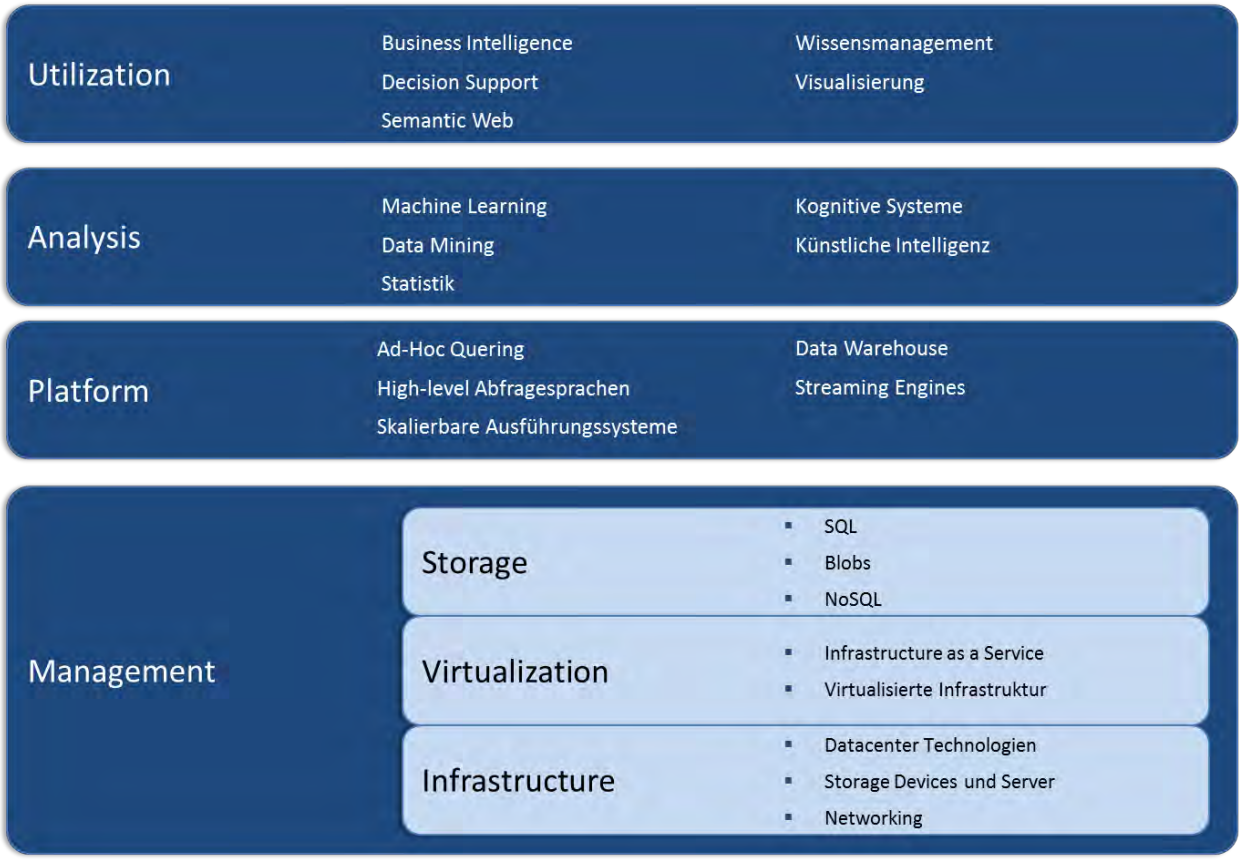


Abbildung 5: Überblick Big Data Technologien

2.2.1 Utilization

Für die Nutzbarmachung von Daten und die daraus resultierende Generierung von Mehrwert kann eine Vielzahl an verfügbaren Technologien und Verfahren angewandt werden. Eine komplette wissenschaftliche Perspektive auf diese kann innerhalb dieser Studie nicht vollständig abgebildet werden. Aus diesem Grund diskutiert diese Studie selektierte Verfahren, welche den AutorInnen im Kontext von Big Data als besonders wichtig erscheinen. Dies beinhaltet Technologien zur einfachen Visualisierung von großen Datenmengen („Visual Analytics“), Methoden zur Daten- und Wissensrepräsentation und zum Wissensmanagement sowie Verfahren in den Bereichen entscheidungsunterstützende Systeme und Business Intelligence.

2.2.1.1 Visualisierung

Für Entscheidungsträger, Analysten und auch für Wissenschaftler ist die Darstellung und Visualisierung von Informationen essenziell. Diese Informationen und die darunter liegenden Daten sind oft nicht vollständig oder/und beinhalten komplexe Wissensstrukturen, welche durch Beziehungen verschiedenster Art miteinander verbunden sein können. Die Visualisierung von solch komplexen Informationen, im speziellen in interaktiver Art und Weise, erfordert hochskalierbare und adaptive Technologien.

Ein Bereich der sich mit dieser Thematik eingehend beschäftigt nennt sich „Visual Analytics“. Diese Domäne kombiniert automatisierte Analysemethoden mit interaktiven Visualisierungen. Dieser Ansatz ermöglicht die Verbindung der Vorzüge der menschlichen Wahrnehmung in der Erkennung von komplexen Mustern mit der maschinellen Verarbeitung der Daten (Keim, Mansmann, Schneidewind, Thomas, & Ziegler, 2008) und ist in Abbildung 6 näher dargestellt.

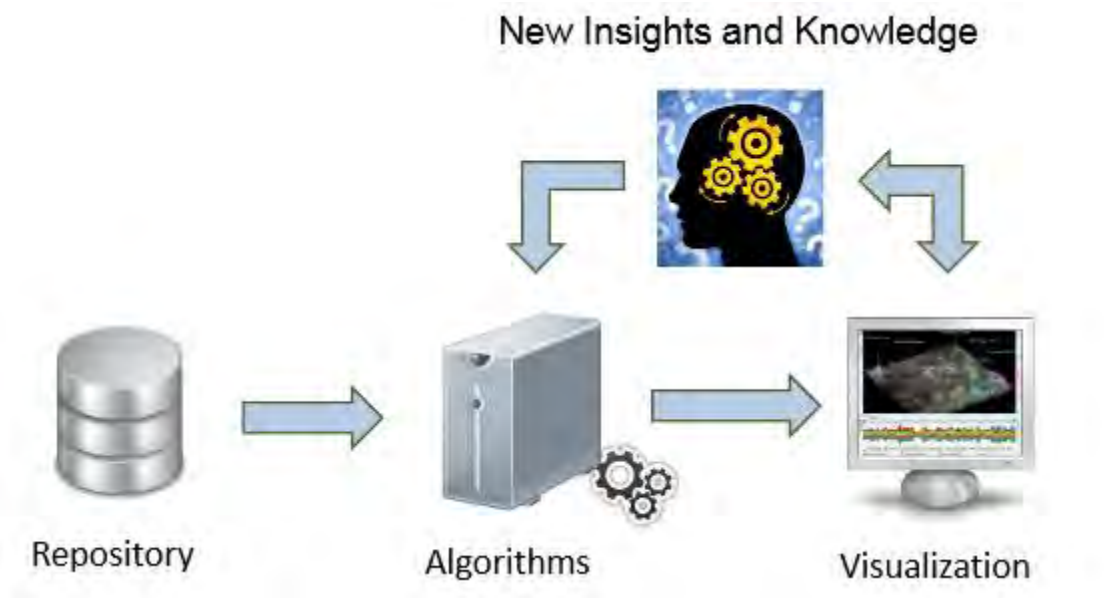


Abbildung 6: Visual Analytics (Sabol, 2013)

Die Problemstellung der effizienten Visualisierung von Daten und Informationen kann des Weiteren in drei Kategorien eingeteilt werden (Sabol, 2013), (Keim, Mansmann, Schneidewind, Thomas, & Ziegler, 2008).

- **Datenvisualisierung** beschäftigt sich mit Daten, denen eine natürliche Repräsentation zugrunde liegt, wie zum Beispiel Karten.
- **Informationsvisualisierung** beschäftigt sich mit der Darstellung von abstrakten Informationen, welche keine natürliche Darstellung besitzen.
- **Wissensvisualisierung** bezeichnet die Darstellung von Wissen, genauer die Kommunikation von explizitem Wissen zwischen menschlichen Personen.

Darüber hinaus kann die Visualisierung von Daten auf unterschiedliche Arten erfolgen. Beispielsweise können Daten als Formalismen dargestellt werden. In diesem Fall werden abstrakte schematische Darstellungen erstellt, welche von Anwendern gelernt und verstanden werden müssen. Eine weitere Methode hier ist die Verwendung von Metaphern. Metaphern erlauben es, Daten auf Basis von Äquivalenten aus der realen Welt darzustellen. Hierbei können Anwender die

Daten auf eine einfache Art und Weise durch die Verwendung von Analogien verstehen. Häufig werden für die Darstellung von Information auch Modelle gewählt, welche ein vereinfachtes und abstraktes Bild der Realität darstellen.

Für strukturierte Daten und Informationen gibt es viele unterschiedliche etablierte Visualisierungen von Linked Data (Berners-Lee, 2001), (Bizer, Heath, & Berners-Lee, 2009) im semantischen Bereich über UML-Diagramme (Bsp: UML (Booch, Rumbaugh, & Jacobson, 1998)), welche in vielen Bereichen als Darstellungsform gewählt werden. Im Zusammenhang mit Big Data stellt sich die Anwendung von Visualisierungsmethoden auf große, unstrukturierte und unterschiedliche Datenquellen als eine große Herausforderung dar. In diesem Zusammenhang müssen neue Ansätze entwickelt bzw. bestehende Ansätze weiterentwickelt werden.

Für die Visualisierung von mehrdimensionalen und komplexen Daten mit unterschiedlichen Beziehungen werden Methoden benötigt, welche mehrdimensionale Skalierung (unterschiedlicher Detaillierungsgrad) und die Aggregation von Daten bei gleichzeitiger Bereitstellung von Interaktivität und Übersichtlichkeit ermöglichen. Technologische Ansätze in diesem Bereich reichen von hocheffizienten JavaScript-Bibliotheken wie zum Beispiel D3[2], Sigma,js[3], oder Arbor.js[4] bis zu Unternehmen, welche sich auf die einfache Visualisierung von aus großen Datenmengen extrahierten Informationen spezialisieren, z.B. MapLarge API[5] oder auch WebLyzard[6].

2.2.1.2 Wissensmanagement und Semantic Web

Wissen wird von vielen Unternehmen als die wichtigste Quelle für nachhaltige Wettbewerbsvorteile gesehen (Trillitzsch, 2004). Aus diesem Grund erhält das Management von Wissen in zahlreichen Unternehmen und in der Forschung einen besonders hohen Stellenwert – stellt sich aber gleichzeitig als große Herausforderung dar. Wissen kann auf unterschiedlichste Arten definiert werden. Als Beispiele werden hier die Erkenntnistheorie (Popper, 2010) sowie die Definition von Wissen im Bereich der künstlichen Intelligenz in (Wahlster, 1977) als Ansammlung von Kenntnissen, Erfahrungen und Problemlösungsmethoden als Hintergrund für komplexe Informationsverarbeitungsprozesse. Wissen und Wissensmanagement bilden einen wichtigen Anknüpfungspunkt zu Big Data Technologien vor allem in Hinsicht auf das Verständnis der Daten und der maschinellen und automatisierten Wissensextraktion.

Semantische Technologien, welche das Ziel haben, Wissen explizit darzustellen, um dieses verarbeitbar zu machen (Berners-Lee, 2001), zählen zu den bedeutesten Initiativen in diesem Bereich. Dafür werden hier oftmals Ontologien verwendet, die eine Basis für die logische Repräsentation und Abfrage von Wissen und für deren Verknüpfung mit zusätzlicher Information zur Verfügung (McGuiness & Van Harmelen, 2004) stellen. Die treibende Kraft hinter der Weiterentwicklung und Definition von semantischen Technologien und Ontologien ist das World Wide Web Consortium (W3C)[7]. Das W3C veröffentlicht regelmäßig neue Empfehlungen und Spezifikationen für diesen Bereich. Dabei wurden unterschiedliche Technologien mit dem Ziel der Wissensrepräsentation erstellt, die wichtigsten hierbei sind Repräsentationssprachen, wie z.B.

[2] D3: http://d3js.org/
[3] Sigma: http://sigmajs.org/
[4] Arbor: http://arborjs.org/
[5] MapLarge: http://maplarge.com/
[6] WebLyzard: http://www.weblyzard.com/
[7] W3C: http://www.w3.org/

DAML+OIL[8], das Resource Description Framework Schema (RDFS)[9] und die Web Ontology Language (OWL)[10], welche zum Quasi-Standard für Ontologien geworden ist. OWL stellt zwei Möglichkeiten zur Verfügung, um Ontologien Bedeutung zuzuordnen: Direct Semantics und RDF Based Semantics. Letztere ermöglicht die Kompatibilität mit dem Resource Description Framework (Graphen-basierte Semantik), während die erste Kompatibilität mit SROIQ-Beschreibungslogik bietet.

Eine der größten Herausforderungen in diesem Bereich ist die enge Kopplung von Wissensmanagement und Repräsentationstools mit Tools für die Analyse und Verarbeitung von riesigen Datenmengen, auf welche in den nächsten Abschnitten näher eingegangen wird. Eine enge Verschränkung dieser Technologien unter Berücksichtigung der Skalierung in Bezug auf Rechenressourcen und Datenmengen kann helfen, das Potenzial beider Technologien voll auszuschöpfen.

LinkedData ist ein weiterer wichtiger Ansatz für die Darstellung und Verlinkung von Wissen (Bizer, Heath, & Berners-Lee, 2009). LinkedData ist ein Paradigma für die Veröffentlichung und Bereitstellung von verlinkten Daten im Web. Für gewöhnlich werden in diesem Fall Daten als RDF Triples dargestellt, was die Erstellung von riesigen Graphen, die Teile vom Web abbilden, ermöglicht. Dieser Ansatz ermöglicht es, neues Wissen automatisiert durch die Traversierung der Links aufzufinden, und im eigenen Kontext zu nutzen. LinkedData fokussiert auf die Bereitstellung und Vernetzung der Daten während der Bereich von Big Data hauptsächlich die Verarbeitung und Analyse von großen Datenmengen beinhaltet. In letzter Zeit zielen zahlreiche Bemühungen in der Wissenschaft und in der Industrie auf die effiziente Verbindung von beiden Technologien ab (z.B. Proceedings of ESWC, 2013).

2.2.1.3 Business Intelligence und Decision Support

Für EntscheidungsträgerInnen ist es wichtig, möglichst aussagekräftige Informationen als Entscheidungsgrundlage in ihrer Organisation zur Verfügung zu haben. Dafür werden Verfahren und Tools benötigt, welche die Bereitstellung von zusätzlichen (und kondensierten) Informationen auf Basis der richtigen Fragestellung ermöglichen. Ziel von Business Intelligence (Grothe & Schäffer, 2012) ist es, effiziente Indikatoren auf Basis von breiten Datenanalysen bereitzustellen. Business Intelligence Tools fokussieren auf die Unternehmensstrategie, Leistungskennzahlen und Reporting und werden unabhängig vom Geschäftsfeld eingesetzt. In diesem Bereich werden zwei Themen als besonders wichtig erachtet: entscheidungsunterstützende Systeme und Trend Analyse.

Entscheidungsunterstützende Systeme betrachten viele Aspekte, angefangen mit Funktionen anspruchsvoller Datenbanksysteme über Modellierungswerkzeuge bis hin zu BenutzerInnenschnittstellen, die interaktive Abfragen ermöglichen (Shim, Warketin, Courtney, Power, Sharda, & Carlsson, 2002). Solche Systeme werden im Forschungs- sowie im Unternehmensumfeld eingesetzt, um menschlichen BenutzerInnen bei der Entscheidungsfindung zu helfen. Dafür werden Informationen aus unterschiedlichen Datenquellen aufbereitet, Modelle darauf angewendet und die Resultate visualisiert.

[8] DAML+OIL: http://www.daml.org/2001/03/daml+oil
[9] RDF Schema 1.1: http://www.w3.org/TR/rdf-schema/
[10] OWL 2 Web Ontology Language: http://www.w3.org/TR/owl2-overview/

In beiden Bereichen gibt es eine Vielzahl internationaler sowie nationaler Anbieter von unterschiedlichen Frameworks für spezielle Anwendungsfälle. In Abschnitt 2.3 werden die wichtigsten Marktteilnehmer in Bezug auf Österreich dargestellt und deren Angebote detaillierter analysiert.

2.2.2 Analytics

Der Bereich Analytics beschäftigt sich mit der Informationsgewinnung aus großen Datenmengen auf Basis von analytischen Ansätzen. Das Ziel hierbei ist es unter anderem, neue Modelle aus Datenmengen zu erkennen, vorhandene Muster wiederzufinden oder neue Muster zu entdecken. Dafür können komplexe analytische Verfahren und Methoden aus unterschiedlichen Bereichen eingesetzt werden, wie z.B. Machine Learning, Data Mining, Statistik und Mathematik sowie kognitive Systeme. Diese Verfahren und Technologien müssen aber in weiterer Folge an die neuen Gegebenheiten im Bereich Big Data (Datenmenge, zeitliche Komponente sowie unterschiedliche und unstrukturierte Daten) angepasst werden beziehungsweise müssen diese teilweise neu entwickelt werden. In diesem Kapitel wird in weiterer Folge auf bestehende Analytics Technologien für Big Data näher eingegangen.

In (Bierig, et al., 2013) wird eine komplexe Analyse von Intelligent Data Analytics in Bezug auf Österreich durchgeführt. Dieser Bereich wird als die Extraktion von Bedeutung aus Daten klassifiziert und in vier interagierende Gruppen eingeteilt: Suche und Analyse, semantische Verarbeitung, kognitive Systeme sowie Visualisierung und Präsentation. Hier wird der Bereich Big Data anhand des Big Data Stacks definiert und die genannten Technologien werden demnach teilweise auf der Ebene von Utilization („die Nutzbarmachung von Daten“) und auf der Ebene der Analytics („analytische Modelle und Algorithmen“) klassifiziert. Semantische Technologien sowie Visualisierung und Präsentation werden in dieser Studie als Utilization-Technologien eingereiht, während kognitive Systeme und Analyse als analytische Probleme klassifiziert werden.

Herausforderungen und Anwendungen im Bereich Big Data Analytics kommen zu einem Teil aus traditionellen High Performance Computing (HPC) Anwendungsfeldern in der Forschung, Industrie und der öffentlichen Hand. Durch die erhöhte Verfügbarkeit und von Big Data Analytics entstehen immer neue Anwendungsfelder von der Produktion (z.B. parametrische Modelle) über Finanzdienstleistungen (z.B. stochastische Modelle) bis hin zu Anwendungen in kleinen und mittleren Unternehmen.

2.2.2.1 Machine Learning

Der Bereich Machine Learning beschäftigt sich mit der Erstellung von Systemen für die Optimierung von Leistungskriterien anhand von Beispieldaten oder vergangenen Erfahrungen (Alpaydin, 2010). Diese Systeme basieren meistens auf Modelle, die mit Hilfe von Machine Learning Algorithmen auf das Verhalten vorhandener Daten „trainiert“ werden. Die Modelle können in deskriptive und preskriptive Modelle unterteilt werden. Deskriptive Modelle versuchen neue Informationen aus vorhandenen Daten zu erlernen und preskriptive Modelle erstellen Vorhersagen für die Zukunft. Des Weiteren kann Machine Learning in überwachtes, unüberwachtes und bestärkendes Lernen unterteilt werden. Methoden für überwachtes Lernen versuchen eine Hypothese aus bekannten (Zielwert und/oder Belohnungswert) Daten zu erlernen, während bei unüberwachtem Lernen Algorithmen angewendet werden, die ohne im Voraus bekannte Zielwerte oder Belohungswerte auskommen. Demgegenüber verwendet bestärkendes Lernen Agenten, welche den Nutzen von konkreten Aktionsfolgen bestimmen.

Unabhängig von dieser Einteilung können im Bereich Machine Learning unterschiedliche Verfahren angewendet werden, um neue Modelle zu lernen oder bestehende Modelle zu verbessern. Viele Verfahren beruhen auf statistischen Modellen, aber es werden auch häufig Algorithmen aus den Bereichen neuronale Netze, Ensemble Learning und genetische Algorithmen verwendet.

Im wissenschaftlichen Bereich stehen viele unterschiedliche Projekte für die Anwendung von Machine Learning zur Verfügung. Das WEKA Toolkit[11] bietet eine Sammlung unterschiedlicher Machine Learning Algorithmen. Dieses umfasst Verfahren für die Vorverarbeitung, Klassifizierung, Regression, Clustering und die Visualisierung der Daten. Gerade für den Einsatz von Big Data ist das Projekt Apache Mahout[12] von immenser Bedeutung, da es sich mit der effizienten und parallelisierten Umsetzung von Machine Learning-Algorithmen beschäftigt und diese auf Open Source-Basis zur Verfügung stellt. Im letzten Jahr hat auch das Apache Spark Projekt[13] mit der MLlib Machine Learning Bibliothek im Bereich Big Data für Aufsehen gesorgt. MLlib bietet ein breites Set an Machine Learning-Algorithmen welche auf einfache Art und Weise auf große Datenmengen angewendet werden können und darunter liegend auf hochperformanten Bibliotheken beruht.

2.2.2.2 Data Mining

Der Bereich Data Mining wird meistens innerhalb des Prozesses für Knowledge Discovery in Databases (KDD) (Fayyad, Piatestsky-Shapiro, & Smyth, 1996) definiert. Der Prozess besteht aus mehreren Schritten, angefangen von der Selektion, der Vorverarbeitung, der Transformation, dem Data Mining und der Interpretation sowie der Evaluation. In dem Schritt Data Mining können unterschiedliche Methoden angewandt werden, welche sich mit der Anwendung von passenden Modellen auf oder dem Auffinden von interessanten Mustern in den beobachteten Daten beschäftigen (Han & Kamber, 2006). Die Anwendung von Modellen kann in zwei Bereiche unterteilt werden: statistische Modelle, welche nichtdeterministische Effekte zulassen sowie deterministische und logische Herangehensweisen (Fayyad, Piatestsky-Shapiro, & Smyth, 1996). Die hierbei verwendeten Verfahren und Methoden basieren meist auf Lösungen aus anderen Bereichen, wie zum Beispiel Machine Learning, Klassifizierung, Clustering und Regression.

Die Begriffe Machine Learning und Data Mining werden oft mit ähnlicher Bedeutung verwendet. Demzufolge kann hier als Unterscheidung zwischen Data Mining und Machine Learning die Anwendung der Algorithmen auf große Datenmengen bzw. auf Datenbanken verstanden werden (Alpaydin, 2010). Des Weiteren werden Machine Learning-Verfahren häufig im Data Mining auf große Datenmengen angewendet.

2.2.2.3 Statistik

Viele Verfahren und Methoden in den Bereichen Data Mining und Machine Learning basieren auf statistischen Methoden. Demzufolge ist ein profundes Verständnis von Statistik für die Anwendung von Data Mining- und Machine Learning-Analysen von Nöten.

[11] WEKA Toolkit: http://www.cs.waikato.ac.nz/ml/weka/
[12] Apache Mahout: http://mahout.apache.org/
[13] Apache Spark: http://spark.apache.org

Ein System, das sehr breite Verwendung findet und auch an die Erfordernisse im Bereich Big Data angepasst wird, ist das Projekt R[14]. R ist eine Open Source-Lösung und bietet sowohl Konnektoren für die effiziente Ausführung auf großen Datenmengen als auch Visualisierungswerkzeuge. Ein weiterer wichtiger Marktteilnehmer in diesem Bereich ist das Unternehmen SAS[15]. SAS bietet unterschiedliche Softwarelösungen für Big Data Analytics.

2.2.2.4 Kognitive Systeme und künstliche Intelligenz

Der Bereich der künstlichen Intelligenz beschäftigt sich seit mehr als 50 Jahren mit Prinzipien, die intelligentes Verhalten in natürlichen oder künstlichen Systemen ermöglichen. Dies beinhaltet mehrere Schritte beginnend bei der Analyse von natürlichen und künstlichen Systemen, der Formulierung und dem Testen von Hypothesen für die Konstruktion von intelligenten Agenten bis zu dem Design und der Entwicklung von computergestützten Systemen, welche aus allgemeiner Sicht Intelligenz erfordern (Poole & Mackworth, 2010).

Manche der bisher genannten Systeme haben breite Schnittmengen mit dem Bereich der künstlichen Intelligenz (Machine Learning sowie Wissensrepräsentation und Reasoning). In (Bierig, et al., 2013) werden prädiktive Techniken in diesen Bereich eingeteilt, welche anhand von Testdaten Verbindungen zwischen Input- und Output-Daten modellieren und auf dieser Basis Vorhersagen für weitere Input-Daten treffen können.

Hier werden unter kognitiven Systemen und künstlicher Intelligenz alle Technologien zusammengefasst, welche sich mit der Transformation von Daten in Wissensstrukturen beschäftigen und auf deren Basis (menschliches) Handeln vorhersehen.

2.2.3 Platform

Die Platform-Ebene beschäftigt sich mit der effizienten Ausführung von Datenanalyseverfahren auf großen Datenmengen. Die Hauptaufgaben hierbei sind die Bereitstellung von parallelisierten und skalierenden Ausführungssystemen sowie die Echtzeitanbindung von Daten, die auf großen und unstrukturierten Datenmengen Anwendung finden. Dabei werden abstrahierten Abfragesprachen zur Verfügung gestellt, die die Verteilung der Daten und die tatsächliche verteilte Ausführung der Abfragen verbergen.

2.2.3.1 Skalierbare Ausführungssysteme

Eine der wichtigsten Innovationen für die breite Anwendung von Big Data Analysen ist das MapReduce-Programmierparadigma (Ekanayake, Pallickara, & Fox, 2008). MapReduce basiert auf einem massiv parallelen Parallelisierungsansatz und verwendet funktionale Programmierkonstrukte ohne Nebeneffekte. MapReduce-Programme werden auf verteilten Datensätzen ausgeführt und unterteilen ein Programm in zwei nebeneffektfreie Funktionen: die Map-Funktion, welche Informationen aus Daten extrahiert, und die Reduce-Funktion welche die Informationen zusammenführt. MapReduce bietet die Möglichkeit, Datenanalysen auf unstrukturierten Daten sehr einfach zu parallelisieren und auszuführen, konzentriert sich aber hauptsächlich auf Probleme, welche mittels Batch-Verarbeitung gelöst werden können, also nicht zeitsensitiv sind. Das wichtigste

[14] The R Project: http://www.r-project.org/
[15] SAS: http://www.sas.com

MapReduce Toolkit ist das Hadoop Framework[16] von der Apache Software Foundation. Ein Schritt in die Richtung einer effizienteren Ausführungsplattform ist den Entwicklern mit der Weiterentwicklung der Plattform zu Apache YARN gelungen. YARN bietet ein effizientes Grundgerüst für die Implementierung von unterschiedlichen parallelisierten Ausführungsumgebungen.

In unterschiedlichen Projekten wird das MapReduce-Paradigma weiterentwickelt bzw. werden auch andere Paradigmen umgesetzt, um die Möglichkeiten der Parallelisierung zu erweitern und zeitsensitive Analysen zu ermöglichen. Ein erster Ansatz für die Erweiterung des Paradigmas ist iteratives MapReduce, welches die Zwischenspeicherung von temporären Resultaten ermöglicht. Des Weiteren wird immer häufiger das Bulk Synchronous-Programmiermodell (BSP) (Valiant, 1990) verwendet. BSP ist ein lange bekannter Ansatz für die Parallelisierung von Programmen und eignet sich hervorragend für datenintensive Aufgaben. In den letzten Jahren wurden mehrere Programmiermodelle für Big Data auf Basis des BSP-Modells erstellt und es stehen erste innovative aber prototypische Frameworks hierfür zur Verfügung. Eine Herangehensweise für Big Data anhand von BSP wurde von Google mit seinem Pregel System (Malewicz, et al., 2010) vorgestellt. Apache Hama bietet die Möglichkeit, BSP basierte Programme auf großen Datenmengen auszuführen, und Apache Giraph[17] ermöglicht iterative Graphenverarbeitung auf dieser Basis.

An dieser Stelle darf auch der klassische HPC-Bereich nicht vergessen werden, der die Grundlage für Innovationen in diesem Bereich bietet. Auf Basis von etablierten HPC-Programmiermodellen wie MPI (Message Passing Interface Forum, 2012), OpenMP (OpenMP Application Program Interface, Version 4.0, 2013) oder auch Sector/Sphere (Gu & Grossman, 2007) können effiziente datenintensive Applikationen implementiert werden und neue Programmiermodelle für Big Data entstehen. Innovationen in diesem Bereich können die zeitsensitive Ausführung von Big Data Analysen maßgeblich voranbringen. Ein weiteres gerade für parallele Programmierung noch nicht gelöstes Problem besteht hierbei darin, dem Benutzer die richtige Mischung aus Abstraktion (einfache Parallelisierungskonstrukte und simple Programmierung) und Effizienz (schnelle Ausführungszeit und Energieeffizienz) zu bieten. Die Lösung für diese Herausforderung ist aber die Grundlage für effiziente und energiesparende Systeme nicht nur im Big Data Bereich.

2.2.3.2 Ad-hoc-Abfragen

Die Grundlage für Echtzeitanwendungen und die interaktive Verwendung von Daten in Anwendungen ist, dass Antworten auf gestellte Anfragen zeitnah geliefert werden können. Für strukturierte Daten liegen unterschiedliche Verfahren und Werkzeuge vor (z.B. relationale Datenbankmanagementsysteme). Wenn strukturierte Daten mit unstrukturierten Daten verknüpft werden, unstrukturierte Daten analysiert werden sollen oder die Datenmengen steigen, kann jedoch mit solchen Systemen die Abfrage von Informationen in Echtzeit ein Problem darstellen. Die Bereitstellung von adäquaten Systemen stellt ein wichtiges Forschungsgebiet dar, in dem immer größere Fortschritte erzielt werden.

Die derzeit wichtigsten Fortschritte in diesem Bereich wurden in Googles Dremel System (Melnik, et al., 2010) umgesetzt, welches die Basis für Google BigQuery[18] bietet. An einer Open Source-Lösung mit dem Namen Drill[19] wird derzeit von der Apache Software Foundation gearbeitet. Zwei weitere

[16] Apache Hadoop: http://hadoop.apache.org/
[17] Apache Giraph: http://giraph.apache.org/
[18] Google BigQuery: https://developers.google.com/bigquery/
[19] Apache Drill: http://incubator.apache.org/drill/

interessante und innovative Initiativen in diesem Bereich sind Clouderas Impala[20] sowie die Stinger Initiative[21], welche von Hortonworks initiiert wurde. Beide Technologien ermöglichen die Beschleunigung von Hadoop-basierten Applikationen in Richtung Echtzeitanalyse. Auch Apache Spark findet in diesem Bereich eine immer höhere Akzeptanz durch einen neuartigen kombinierten Ansatz aus paralleler Programmierung, in-memory und simplen high-level APIs. Generell kann hier ein Trend zu In-Memory-Systemen beziehungsweise zur Kombination von In-Memory und verteilten Speichersystemen beobachtet werden.

2.2.3.3 High-Level-Abfragesprachen

Unabhängig von den verwendeten Systemen für die Speicherung der Daten und für die Ausführung von Algorithmen wird eine abstrahierte und standardisierte Möglichkeit der Datenabfrage benötigt. Während sich im Bereich der relationalen Datenbankmanagementsysteme seit langem SQL als Standard durchgesetzt hat, werden gerade im Big Data Bereich oft andere und simplere (auf die Funktionalität bezogen) Abfragemöglichkeiten geboten. Gerade auch in Hinblick auf die parallelisierte Ausführung der Abfragen werden in diesem Bereich Verfahren und Technologien adressiert, welche einen simplen Zugriff auf große Datenmengen und im besten Fall auch der Analysen ermöglichen. Generell kann das Ziel dieser Ebene als das gleichzeitige Anbieten der Komplexität und des Funktionsumfanges von SQL und von darunter liegenden massiv parallelen Programmiermodellen definiert werden. Derzeit befinden sich die meisten Systeme zwischen dem Anbieten der kompletten Funktionalität von SQL und der massiven Skalierbarkeit, sind demzufolge entweder in die eine oder andere Richtung spezialisiert. Hier ist aber ein Angleichen der Funktionalitäten der unterschiedlichen Systeme zu erkennen.

Wichtige Systeme in diesem Bereich, welche seit einigen Jahren auf breiter Basis von Firmen eingesetzt und vermarktet werden, sind vor allem Apache Hive[22] und Apache Pig[23]. Während Apache Hive eine Data Warehouse-Ebene über die darunter liegenden Daten und Ausführungsebenen legt und damit eine abstrahierte Zugriffsschicht bietet, verfolgt Pig den Ansatz einer simplen und abstrahierten Scripting-Sprache für die Analyse von großen Datenmengen.

2.2.3.4 Data Warehouse

Der Bereich Data Warehouse spielt ebenfalls eine tragende Rolle für Big Data. Data Warehouses sind etablierte Technologien, welche in den letzten Jahren an die neuen Herausforderungen von Big Data angepasst wurden. Diese werden oft zur Konsolidierung von unterschiedlichen Datenquellen und für deren unternehmensinterne Auswertung (z.B. Business Intelligence, Reporting, …) verwendet. Neben Systemen, die landläufig als „Big Data Systeme“ verstanden werden (NoSQL, MapReduce, etc.), haben diese einen bedeutenden Einfluss auf die Datenverwaltung und deren Management. Immer häufiger werden diese mit Big Data Ansätzen ergänzt, um unstrukturierte und Echtzeitdaten zu integrieren. Der wesentliche Unterschied zu den bisher beschriebenen Big Data Systemen besteht in der Herangehensweise bei der Auswahl der gespeicherten Daten. Während das Ziel bei Big Data Systemen ist, alle Rohdaten zu speichern und für zukünftige Analysen bereitzuhalten, fokussieren sich Data Warehouses auf Sets an vorextrahierten Daten, welche für spezifische Anwendungsfelder benötigt werden.

---

[20] Cloudera Impala: http://www.cloudera.com/content/cloudera/en/products-and-services/cdh/impala.html
[21] Stinger Initiative: http://hortonworks.com/labs/stinger/
[22] Apache Hive: http://hive.apache.org/
[23] Apache Pig: http://pig.apache.org/

2.2.3.5 Streaming Engines

In einer Vielzahl von Bereichen stehen immer mehr permanent Daten liefernde Sensoren zur Verfügung. Daraus entsteht ein erhöhter Bedarf nach Verfahren und Technologien für deren Verarbeitung in Echtzeit. Ein prominentes Beispiel dazu liefert der Teilchenbeschleuniger am CERN (LHC) (Wadenstein, 2008), deren Echtzeit-Datenumfang Dimensionen annimmt die eine komplette Speicherung oftmals unmöglich macht. In solchen Systemen müssen Daten direkt nach ihrer Erzeugung verdichtet werden, um möglichst viele Informationen zu erhalten und diese Daten für die Weiterverarbeitung nutzbar zu machen.

Hier wird auf zwei relevante Open Source-Projekte für Streamverarbeitung hingewiesen. Das Apache Projekt S4[24] ermöglicht die parallele Verarbeitung von Datenstreams von verteilten Ressourcen und bietet Fehlertoleranz. Das Apache Projekt Storm[25] bietet ebenfalls eine verteilte und fehlertolerante Ausführungsumgebung für Streamverarbeitung und für Echtzeitanalysen. Erwähnenswert ist auch die direkte Integration dieser Tools mit dem Apache Hadoop Projekt was eine Speicherung und weiterführende Analyse der Daten mit zum Beispiel MapReduce ermöglicht. Apache Storm wird unter anderem von Twitter und Groupon für die Echtzeitverarbeitung von Streams verwendet.

2.2.4 Management

Die Verarbeitung von großen und unterschiedlich strukturierten Datenmengen erfordert innovative Ansätze für deren Verwaltung. Erstens werden Systeme für die effiziente und verteilte Speicherung von den Daten benötigt. In diesen Bereich fällt neben den klassischen relationalen Datenbankmanagementsystemen vor allem der Bereich der NoSQL-Systeme. Zusätzlich sind verteilte Dateisysteme und Binary Object Stores von Bedeutung. Zweitens benötigen diese Systeme für ihre Ausführung effiziente Recheninfrastrukturen. In dieser Studie werden Infrastrukturthemen in die Virtualisierungs- und Infrastrukturthematik eingeteilt. Virtualisierung beschäftigt sich mit der transparenten Bereitstellung von Ressourcen und umfasst in dieser Studie Cloud Computing und Hardwarevirtualisierungslösungen. Auf der Infrastrukturebene wird die Bereitstellung von Datenzentren und HPC-Infrastruktur beleuchtet.

2.2.4.1 Speicherlösungen

Relationale Datenbanksysteme (RDBMS) folgen einem wohldefinierten und strikten Schema und erfüllen die ACID Prinzipien (Atomarität, Konsistenzerhaltung, Isolation, Dauerhaftigkeit) (Haerder & Reuter, 1983). Darüber hinaus bieten relationale DBMS eine flexible und standardisierte Abfragesprache (SQL), welche komplexe Abfragen der Daten zulässt. Diese Systeme bieten somit strikte Konsistenz (alle BenutzerInnen sehen zu jeder Zeit dieselben Daten) und alle Daten stehen auch jederzeit für jeden Benutzer zur Verfügung (Verfügbarkeit). Für die Skalierung eines relationalen DBMS stehen unterschiedliche Methoden zur Verfügung, wie z.B. Sharding, Partitioning oder eine Master/Slave-Architektur. Wenngleich aktuelle relationale DBMS die Speicherung und Verwaltung von größeren Datenmengen durch diese Verfahren unterstützen, sind diese Systeme nicht auf eine große Anzahl von Rechnern skalierbar und damit auch in der effizienten Verarbeitung von großen Datenmengen begrenzt. Diese Problemstellung wurde von Eric Brewer im Jahr 2000 im CAP Theorem (Brewer, Towards Robust Distributed Systems, 2000) genauer beschrieben. Das

[24] Apache S4: http://incubator.apache.org/s4/
[25] Apache Storm: https://storm.incubator.apache.org

Theorem beschreibt die wesentlichen Kriterien eines verteilten Datenbanksystems als Konsistenz (ACID), Verfügbarkeit und Partitionstoleranz (Skalierbarkeit). Das Theorem beschreibt, dass es für ein DBMS nicht möglich ist, alle drei Charakteristiken zu unterstützen. Diese Aussage wurde mathematisch bewiesen, wenngleich keine dieser Charakteristiken als binär zu verstehen ist (Brewer, Pushing the CAP: Strategies for Consistency and Availability, 2012). Aufgrund spezifischer Anforderungen werden immer mehr unterschiedliche DBMS entwickelt, welche sich auf andere Charakteristika spezialisieren und dafür entweder keine strikte Konsistenz bieten oder die durchgängige Verfügbarkeit der Daten nicht garantieren. Hieraus ist der Begriff BASE entstanden, welcher für „basically available, soft state, eventually consistency" steht. BASE-Systeme sind schlussendlich konsistent, versprechen aber nicht, dass jede/r BenutzerIn zu jeder Zeit dieselbe Sicht auf die Daten hat.

Viele in diesem Bereich entstandenen Datenbanken werden unter dem Begriff NoSQL (Edlich, Friedland, Hampe, Brauer, & Brückner, 2011), (Leavitt, 2010) zusammengefasst. NoSQL steht dabei für „nicht-nur SQL" und meint somit, dass es für bestimmte Einsatzszenarien andere besser geeignete Systeme als SQL-basierte Systeme gibt. NoSQL-Systeme können in Key/Value-Systeme, spaltenorientierte Datenbanken, dokumentenbasierte Datenbanken und graphenbasierte Datenbanken unterteilt werden.

Spaltenorientierte Datenbanken speichern alle Werte einer Spalte in einer Datei (im Gegensatz zu den meisten relationalen DBMS). Dieses Ablageformat hat für spaltenorientierte Analysen einen beachtlichen Geschwindigkeitsvorteil. Exemplarische und wichtige Vertreter (in Bezug auf Big Data) in diesem Bereich sind Apache HBase[26] (wide-column store) und Apache Cassandra[27] (Mischung aus spaltenorientiertem und Key/Value-System).

Key/Value-Systeme bauen meist auf verteilten Implementierungen von Hashfunktionen oder B+-Bäumen auf. Auf deren Basis ermöglichen sie effizienten Zugriff auf einzelne Werte, bieten aber meist nur eine simplifizierte Benutzerschnittstelle. Durch diese Implementierung ermöglichen sie die Skalierung auf eine Vielzahl von Rechenressourcen. Beispielumsetzungen von Key/Value-Systemen sind BerkeleyDB[28], REDIS[29], Amazon Dynamo[30] und Chordless[31].

Ein immer wichtiger werdender Bereich ist die effiziente Speicherung und Verarbeitung von Graphen. Durch die weite Verbreitung von sozialen Netzwerken, dem semantischen Web und auch von GIS-Systemen werden immer mehr Daten in einer Form gespeichert, in der die Relationen zwischen den einzelnen Datensätzen mindestens genauso wichtig sind wie die Daten selbst. Dies ermöglicht die Traversierung der Datensätze über deren Beziehungen sowie die Anwendung von graphenbasierten Algorithmen. Dies führt zur Umsetzung von effizienten graphenbasierten Datenbanken wie zum Beispiel Hypergraph[32], Dex[33], Infogrid[34] und VertexDB[35], welche ein breites Angebot an Technologien bieten.

[26] Apache HBase: http://hbase.apache.org/
[27] Apache Cassandra: http://cassandra.apache.org/
[28] BerkeleyDB: http://www.oracle.com/technetwork/database/database-technologies/berkeleydb/overview/index.html
[29] REDIS: http://redis.io/
[30] Amazon Dynamo: http://aws.amazon.com/de/dynamodb/
[31] Chordless: http://sourceforge.net/projects/chordless/
[32] Hypergraph: http://hypergraph.sourceforge.net/

Dokumentenorientierte Datenbanken werden vermehrt für Web Applikationen eingesetzt. Diese sind auf die Verwaltung von großen meist JSON-basierten Datensätzen spezialisiert und ermöglichen die Interpretation des Schemas in den Applikationen selbst. DBMS wie Lotus Notes, MongoDB[36] und CouchDB[37] sind wichtige Repräsentanten dieses Bereichs.

Eine weitere wichtige Technologie für die effiziente Speicherung von großen Datensätzen auf verteilten Rechenressourcen sind verteilte Dateisysteme. Das Hadoop Distributed File System (HDFS) wird innerhalb des Apache Hadoop Frameworks als Grundlage für die Daten-lokale Ausführung von Applikationen verwendet. Viele weitere Umsetzungen wie zum Beispiel Lustre[38] und Ceph[39] ermöglichen die performante und verteilte Speicherung auf Basis von Dateien.

2.2.4.2 Infrastruktur und Virtualisierung

Um riesige Datenmengen effizient speichern und verarbeiten zu können, werden Hardwareressourcen von hochskalierbaren Speichersystemen bis zu Rechenressourcen für Datenzentren und HPC-Systemen benötigt. Hierfür wird in drei Teilbereiche unterschieden: Cloud- und Grid-Lösungen, Datenzentren und HPC-Systeme.

Cloud-Lösungen (Buyya, Broberg, & Goscinski, 2011) bieten dem Benutzer die Illusion von virtuell unendlich vielen verfügbaren Rechen- sowie Speicherressourcen und bieten demnach eine einfache Akquise von diesen für Unternehmen und auch die Forschungslandschaft. Cloud-Lösungen verstecken die Details der angebotenen Hardware und basieren auf Technologien für die Umsetzung von großen Datenzentren. Mehrere Unternehmen bieten in Österreich Cloud-Dienste an und es entstehen erste lokale Anbieter von Cloud-Plattformen (Überblick in Kapitel 2.3). Der Forschungslandschaft steht in Österreich derzeit keine gemeinsame Infrastruktur zur Verfügung, sondern es werden institutionsinterne Lösungen verwendet.

Datenzentren werden für den Aufbau von Cloud-Infrastrukturen sowie unternehmensintern für die Bereitstellung von Rechen- und Speicherressourcen benötigt. Für Datenzentren wird meist Commodity Hardware verwendet, um kostengünstig horizontal skalieren zu können.

Im Bereich High Performance Computing (HPC) (Dowd, Severance, & Loukides, 1998) wurde mit dem gemeinsamen Bau des Vienna Scientific Clusters (VSC-I und II)[40] eine Plattform für die TU Wien, die Universität Wien, die Universität für Bodenkultur, die Technische Universität Graz und die Universität Innsbruck geschaffen. Des Weiteren ist Österreich bei PRACE[41], der Partnership for Advanced Computing in Europe, durch die Johannes Kepler Universität Linz vertreten. Dadurch stehen in einem Review-Prozess ausgewählten Projekten europäische HPC-Infrastrukturen zur Verfügung.

---

[33] Dex: http://www.sparsity-technologies.com/dex.php
[34] Infogrid: http://infogrid.org/trac/
[35] VertexDB: https://github.com/stevedekorte/vertexdb
[36] MongoDB: http://www.mongodb.org/
[37] CouchDB: http://couchdb.apache.org/
[38] Lustre: http://wiki.lustre.org/index.php/Main_Page
[39] Ceph: http://ceph.com/
[40] Vienna Scientific Cluster: http://vsc.ac.at/
[41] PRACE: http://www.prace-ri.eu

Für die Umsetzung von Big Data getriebener Forschung und Innovation benötigen Unternehmen wie Forschungseinrichtungen Zugang zu Datenzentren, HPC-Infrastrukturen und/oder Cloud-Lösungen. In der Bereitstellung solcher Infrastruktur auf nationaler Ebene liegt großes Potenzial für Big Data getriebene Forschung und Unternehmen sowie würde diese die verstärkte Teilnahme an internationalen Forschungsprojekten und den Aufbau von technologiegetriebenen Unternehmen ermöglichen.

2.2.5 Überblick über Verfahren und Methoden

Verfahren und Technologien im Bereich Big Data werden innerhalb dieser Studie in Utilization, Analytics, Platform und Management eingeteilt. Jede Ebene beinhaltet eigene Forschungsbereiche und es ist eine Vielzahl an unterschiedlichen Tools und Technologien verfügbar, welche eingesetzt werden kann. In Tabelle 1 wird ein Überblick über Umsetzungen im Kontext von Big Data dargestellt. Es wurden Technologien ausgewählt, die gerade für die Verarbeitung von großen Datenmengen ausgelegt sind bzw. in diese Richtung erweitert werden. Aufgrund der Komplexität dieses Themenfeldes und der großen Schnittmengen zwischen vielen unterschiedlichen wissenschaftlichen und wirtschaftlichen Themenbereichen erhebt die Aufstellung keinen Anspruch auf Vollständigkeit, sondern bietet vielmehr einen Überblick über selektierte und innovative Technologien für den Bereich Big Data.

Tool	Ebene	Approach	Beschreibung	Version
D3	***Utilization Visualisierung***	JavaScript-basierte Visualisierung	Datengetriebene Dokumente, Web Standard-basiert, skalierbare JavaScript-Visualisierungskomponente für HTML, SVG, und CSS	D3.v3 Stabile Version Open Source
Sigma,js	***Utilization Visualisierung***	JavaScript-basierte Visualisierung	Leichtgewichtige JavaScript-Bibliothek auf Basis des HTML Canvas Elements, interaktive Grafiken, Gephi Integration	Version: v0.1 MIT License Open Source
Arbor.js	***Utilization Visualisierung***	JavaScript-basierte Visualisierung	Visualisierung von Graphen auf Basis von Web Workern und jQuery	MIT License Open Source
MapLarge API	***Utilization Visualisierung***	Kommerzielles Visualisierungtool (Kartenmaterial)	Plattform für die Visualisierung und Analyse von großen Datenmengen anhand von Kartenmaterial, kommerzielle Software, Upload von Daten	Cloud-basiert Kommerziell
WebLyzard	***Utilization Visualisierung***	Web Intelligence und Medienbeobachtung	Automatisierte Analyse von Online-Medien, Bereitstellung von Indikatoren für die strategische Positionierung, Trendanalysen	Kommerziell
Protégé	***Utilization Wissensmanagement und semantische Technologien***	Ontologieeditor, Framework für Wissensbasen	Modellierung und Erstellung von Ontologien, Web und Desktop Client, Plug In Support, Support von Reasoning	WebProtégé build 110 Protégé 3.5 Open Source Mozilla Public License
D2RQ	***Utilization Wissensmanagement und semantische Technologien***	LinkedData	Plattform, um relationale Daten als virtuelle RDF-Graphen bereitzustellen, Abfragen mittels SPARQL, Support der Jena API	Version: 0.8.2 Open Source Apache License, Version 2.0
WEKA	***Analytics Machine Learning***	Sammlung von Machine Learning-Algorithmen	Java-basierte Implementierungen von Machine Learning-Algorithmen, Tools für die Vorverarbeitung, Klassifizierung, Regression, Clustering, Visualisierung; erweiterbar	Version: Weka 3 Open Source GNU GPL

Apache Mahout	***Analytics*** ***Machine Learning***	Skalierbare Sammlung von Machine-Learning-Algorithmen	Implementierungen für Klassifizierung, Clustering, kollaboratives Filtern; Implementierungen auf Basis von Apache Hadoop Framework; hoch optimierte Core Libraries	Version: 0.8 Apache Projekt Open Source
DISPEL	***Analytics*** ***Data Mining***	Scripting-Sprache für Data Mining Workflows	Basiert auf Data Flow-Graphen, übersetzt in ausführbare Workflows, entwickelt innerhalb des EU-Projekts ADMIRE	Admire Software Stack Open Source Apache License
R	***Analytics*** ***Statistics***	Statistische Berechnungen und Grafiken	R ist eine Programmiersprache und Ausführungsumgebung für statistische Berechnungen; lineare und nichtlineare Modellierung, Zeitserienanalyse, Klassifizierung, Clustering, …	Version: R 3.0.2 GNU GPL Open Source
SAS	***Analytics***	Business Intelligence, (Visual und High-Performance)Analytics	Weltweiter Anbieter von Business Analytics-Software, unterschiedliche Toolchains	Kommerziell
Apache Spark – MLlib	***Analytics*** ***Machine Learning***	Skalierbare Sammlung von Machine-Learning-Algorithmen	Implementierungen für K-Means Clustering, lineare und logistische Regression, Naive Bayes, In-Memory und Hadoop-basierte (massiv parallele) Ausführung	Version: 0.91 Apache Projekt Open Source
Apache Hadoop	***Platform*** ***Skalierbare Ausführungssysteme***	MapReduce	Datenlokalität, Fehlertoleranz, Map- und Reduce-Funktionen	Apache Hadoop 1.2.1 Stabile Version Open Source
YARN	***Platform*** ***Skalierbare Ausführungssysteme***	Generic	Generisches Ausführungsmodell, detailliertes Ressourcen-Scheduling (z.B. CPU, Speicher), Ausführungsframework für MapReduce 2	Apache Hadoop 2.2.0 Stabile Version Open Source
Twister	***Platform*** ***Skalierbare Ausführungssysteme***	Iterative MapReduce	Statische und dynamische Daten, Support für Iterationen in MapReduce (aufeinanderfolgende MapReduce-Jobs)	Twister v0.9 Open Source
Apache Hama	***Platform*** ***Skalierbare Ausführungssysteme***	Bulk Synchronous Programming	Supersteps und Barrier-Synchronisation, wissenschaftliche Applikationen auf der Basis von HDFS	Apache Hama 0.6.2 Open Source

Pregel	***Platform Skalierbare Ausführungssysteme***	Bulk Synchronous Programming	Supersteps und Barrier-Synchronisation, Graphenverarbeitung	Google internal
Apache Giraph	***Platform Skalierbare Ausführungssysteme***	Bulk Synchronous Programming	Iterative Verarbeitung von Graphen, Pregel-Modell	Apache Giraph 1.0.0 Open Source
Apache S4	***Platform Skalierbare Ausführungssysteme***	Stream processing	Verteilte Stream Computing-Plattform, continuous unbounded data streams, Fehlertoleranz	Incubator status Apache S4 0.6.0 Open Source
Apache Storm	***Platform Skalierbare Ausführungssysteme***	Stream processing	Verteilte Echtzeitverarbeitung von Streams, programmiersprachenunabhängig	Incubator status Apache Storm 0.9.1 Open Source
Hive	***Platform Data Warehouse/Ad-hoc-Abfragen***	Data warehouse/ Ad-hoc query system	High-Level-Abfragesprache, strukturierte Daten, auf Basis von Apache Hadoop	Apache Hive 0.11.0 Open Source
Pig	***Platform Ad-hoc-Abfragen***	High-level Daten Analyse Sprache	Einfache Skriptsprache (Pig Latin), Optimierung, auf Basis von Apache Hadoop, Erweiterbarkeit (benutzerdefinierte Funktionen)	Apache Pig 0.11.1 Open Source
Dremel	***Platform Ad-hoc-Abfragen***	Ad-hoc-Abfragesystem	Nested-Column-Oriented-Data-Modell, SQL-ähnliche Abfragsprache, multi-level-serving trees (Abfrageoptimierung)	Google internal, Bereitgestellt als Cloud Service via Google BigQuery
Apache Drill	***Platform Ad-hoc-Abfragen***	Ad-hoc-Abfragesystem	Open Source-Version von Dremel	Incubator status Open Source
Stinger Initiative	***Platform Ad-hoc-Abfragen***	Ad-hoc-Abfragesystem	Schnelle interaktive Abfragen auf Basis von Hive, Column store, Vector Query Engine, Query Planner	Project geleitet von Hortonworks Open Source
Cloudera Impala	***Platform Ad-hoc-Abfragen***	Ad-hoc-Abfragesystem	Query Execution Engine (parallel databases), auf Basis von HDFS und HBase	Cloudera Impala 1.1.0 Open Source

Enterprise Control Language	***Platform Ad-hoc-Abfragen***	Ad-hoc-Abfragesystem	Data Refinery Cluster, Query Cluster, Dataflow orientierte Sprache, Umsetzung von Datenrepräsentation und von Algorithmen, Wiederverwendung von Ergebnissen	HPCC Systems (LexisNexis) Freie Community Edition Enterprise Edition
Apache Tez	***Platform Skalierbare Ausführungssysteme***	Graph-processing	Komplexe gerichtete azyklische Graphen, auf Basis von YARN	Incubator status Open Source
OGSA-DAI	***Platform Skalierbare Ausführungssysteme***	Data access and integration	Verteilter Datenzugriff, Datenintegration und Datenmanagement, relationale, objektorientierte XML-Datenbanken und Dateien	OGSA-DAI 4.2 Open Source
Stratosphere	***Platform High-level Query Languages***	High-level data analysis language	Erweiterbare High-Level-Skriptsprache, PACT paralleles Programmiermodell, benutzerdefinierte Funktionen, Optimierung, massiv parallele Ausführung	Stratosphere 0.2 Open Source
BerkeleyDB	***Management Storage***	Key-value stores	Bytearrays als Schlüssel und Werte, embedded Datenbank, nebenläufig, Transaktionen, Replikation	Oracle Berkeley DB 12c Berkeley DB Java Edition 5.0.84 Berkeley DB XML 2.5.16 Dual licensing
Chordless	***Management Storage***	Key-value stores	Implementierung von Chord (peer-to-peer) und verteiltem Hashing, Transaktionen, verteilt, skalierbar, redundant	Latest Version 2009-10-01 Open Source
REDIS	***Management Storage***	Key-value stores	Strings, Hashes, Lists, Sets und Sorted Lists werden als Schlüssel unterstützt, In-Memory Daten, Master-Slave-Replikation	Redis 2.6.14 Open Source
Amazon Dynamo	***Management Storage***	Key-value stores	Starke Konsistenz (read), Replikation, Verwendung von Solid State Drives, Fehlertoleranz, Integration von MapReduce	Amazon Cloud Service (Beta)
HBase	***Management Storage***	Column-oriented database	Echtzeit-read/write-Zugriff, verwendet HDFS, Column Families und Regionen, Replikation, binäres Datenformat	Apache HBase 0.94 Open Source

Cassandra	***Management Storage***	Hybrid column-oriented database and key-value stores	Schlussendliche Konsistenz, Keyspaces, Replikation, Column Families (geordnete Sammlung von Reihen)	Apache Cassandra 1.2.8 Open Source
Dex	***Management Storage***	Graph database	Benannte azyklische gerichtete Graphen, ACID, Nebenläufigkeit, horizontale Skalierung, Verfügbarkeit, shortest path, Export Formate (graphml, graphviz, ygraphml)	Dex 4.8 Developed by Sparsity Dex evaluation, research, development – frei kommerzielle Version
HypergraphDB	***Management Storage***	Graph database	Gerichtete Hypergraphen, P2P framework für Datenverteilung, zugrunde liegende Key/Value-Speicherung, Graphenmodellierungsebene (RDF, OWL 2.0, WordNet, TopicMaps,…)	HypergraphDB 1.2 Open Source
MongoDB	***Management Storage***	Document-oriented database	Schema-free, replication, availability, shards automatically, full index support, JSOBN support	MongoDB 2.4.6 Open Source
CouchDB	***Management Storage***	Document-oriented database	Scalable, JSON documents, B-trees, Master-Master setup, HTTP	Apache CouchDB 1.3.1 Open Source
Ceph	***Management Storage/ Virtualisierung***	Verteilter Object Store und verteiltes Dateisystem	Verteilter Object Store und Data System, hochperformant, Zuverlässigkeit, Skalierbarkeit, Support von Apache Hadoop, RESTful API	Version: v0.73 Open Source LGPL2
OpenStack	***Management Infrastructure***	Cloud Toolkit, Infrastructure as a Service	Bereitstellung einer virtuellen Infrastruktur, Rechenspeicher und Netzwerkressourcen	OpenStack Havanna Open Source

Tabelle 1: Überblick über ausgewählte Big Data Technologien

2.3 Erhebung und Analyse der österreichischen MarktteilnehmerInnen

In diesem Abschnitt wird dargestellt, welche Unternehmen in Österreich am Markt aktiv sind und welche Produkte, kategorisiert auf Basis des Big Data Stacks, anbieten. Die Erhebung dieser Daten fand hierbei auf mehreren Ebenen statt: zu Beginn wurde die interne Datenbasis der IDC erhoben. Hierbei wurde eine Liste erstellt, welche eine Vielzahl an Unternehmen enthält, welche im Big Data Umfeld tätig sind. Diese Liste wurde aus mehreren IDC Reports abgeleitet. Die Liste wurde durch klassischen Desktop-Research erweitert, um auch jene Unternehmen aufzufinden, welche in keinen aktuellen Reports abgebildet wurden. Im letzten Schritt wurde ein Kontakt mit jedem Unternehmen aufgenommen und evaluiert in wie fern der österreichische Markt relevant ist.

2.3.1 Marktanalyse

Die folgende Tabelle stellt eine Übersicht der verfügbaren Lösungen im Big Data Bereich dar. Die einzelnen Unternehmen sind in der darauffolgenden Tabelle genauer beschrieben.

	Utilization	Analytics	Platform	Management
Adobe				
ATOS				
Attachmate Group				
BMC				
Braintribe				
Catalysts				
CommVault				
Compuware				
CSC				
Dassault Systemes				
Dell				
Dropbox				
EMC				
Fabasoft Group				
Fujitsu				
Google Inc.				
HDS				
HP				
IBM				
Infor				
Kofax				
Linemetrics				
Microsoft				
MicroStrategy				
mindbreeze				

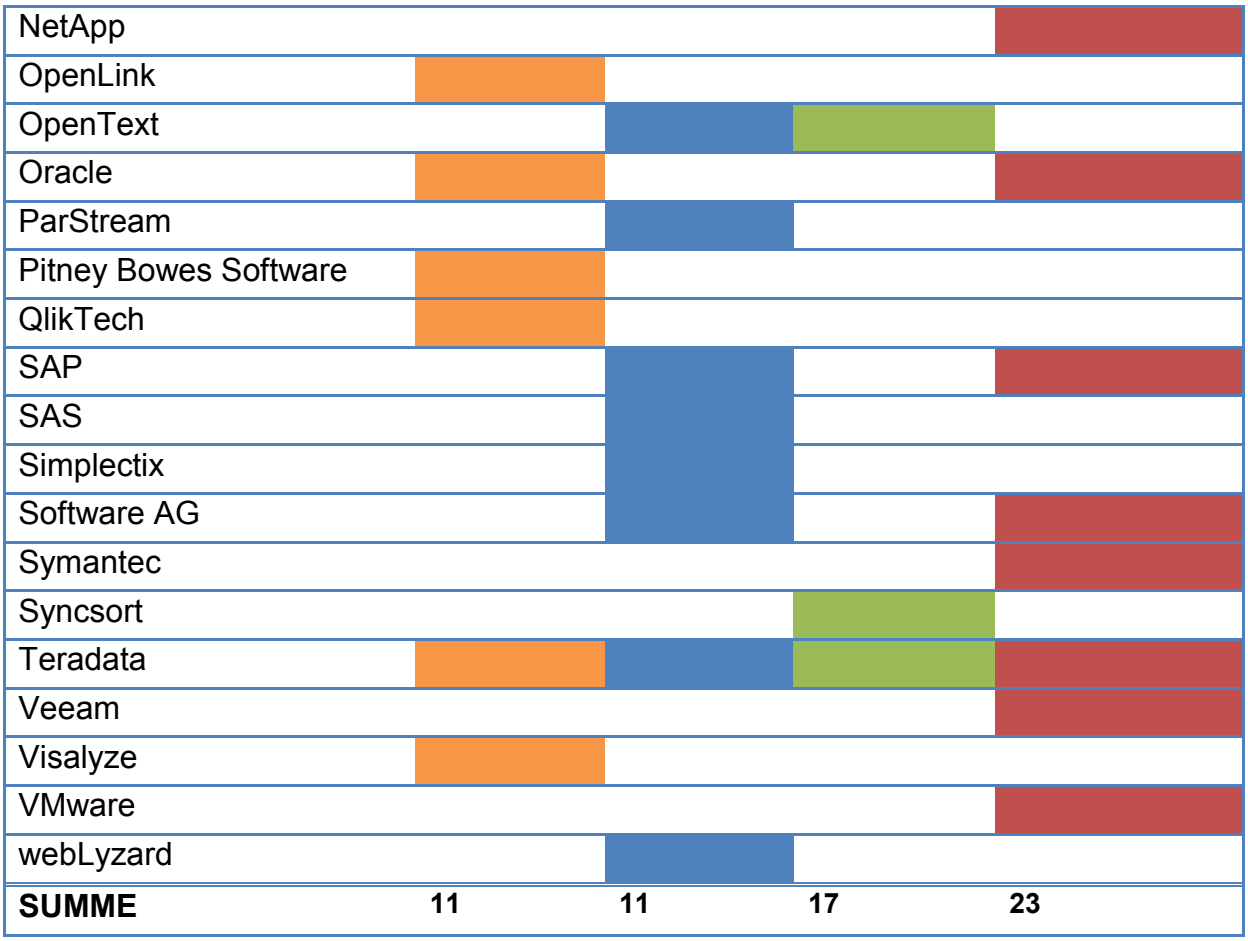


Tabelle 2: Übersicht der Lösungen

In der folgenden Tabelle werden die einzelnen Unternehmen sowohl anhand des Unternehmensprofils als auch anhand der einzelnen Big Data Ebenen beschrieben. Diese tabellarische Aufstellung ist alphabetisch geordnet. Es wurden sämtliche Unternehmen integriert, welche einen Sitz in Österreich haben. Unternehmen, welche keinen Sitz in Österreich haben, jedoch auch von Bedeutung sind, werden in einer eigenen Tabelle aufgelistet.

Adobe

Adobe ist ein US-amerikanisches Softwareunternehmen. Es wurde 1982 gegründet und verfügt über mehr als 9000 Mitarbeiter weltweit bei einem Umsatz von ca. 3,8 Mrd USD. In Österreich selbst besteht eine Zweigniederlassung der deutschen Niederlassung. Adobe ist ein Anbieter für Lösungen im Bereich digitales Marketing und digitale Medien. Betreffend Big Data bietet Adobe Lösungen zur Datenerfassung, Analyse und Visualierung an.

Utilization	-
Analytics	-
Platform	Adobe bietet ein Produkt Namens „Adobe Marketing Cloud“ an. Hierbei handelt es sich primär um Produkt für Markteinblicke und Kundenverhalten. Die Adobe Marketing Cloud bietet Ad-Hoc Abfragen über verschiedenste Elemente. Die Adobe Marketing Cloud richtet

sich an Websitebenutzer. Adobe liefert hier ein vollständiges Service aus, welches in der Cloud gehostet wird. Die Adobe Marketing Cloud besteht aus folgenden Produkten:

- *Adobe Analytics* ist eine generelle Anwendung, welche Informationen über Kunden sammelt und auswertbar macht.
- *Adobe Campaign* ermöglicht eine personalisierte Kommunikation mit dem Kunden und somit eine besseren Einblick darauf, was Kunden sich wünschen.
- *Adobe Experience Manager* erlaubt es, Inhalte auf allen beliebigen Geräten zur Verfügung zu stellen. Damit ist ein einheitliches Erlebnis auf Mobiltelefon, Tablet und PC möglich.
- *Adobe Media Optimizer* bietet einen Überblick darüber, welche Werbeformen wie gut ankommen. Damit kann man die Werbeformen ideal anpassen.
- *Adobe Social* bietet die Möglichkeit, soziale Netzwerke in die Analyse miteinzubeziehen. Hierbei werden vor allem Interaktionen in sozialen Netzen mit Geschäftsergebnissen verglichen.
- *Adobe Target* bietet Lösungen für Tests mit Kunden und wie diese auf Ereignisse in der Website reagieren. Damit soll es einfacher möglich sein, Versuche durchzuführen.

Management -

ATOS

Atos ist ein internationaler Anbieter von IT-Dienstleistungen mit einem Jahresumsatz von 8,5 Milliarden Euro und 76.400 Mitarbeitern in 47 Ländern. Das umfangreiche Portfolio umfasst unter anderem transaktionsbasierte Hightech-Services, Beratung, Systemintegration und Outsourcing-Services. Atos Österreich beschäftigt 1.840 MitarbeiterInnen. Atos ist ein Dienstleistungsunternehmen, welches zwar keine Produkte, jedoch Dienstleistungen für das Thema Big Data anbietet.

Utilization -

Analytics -

Platform Zum Dienstleistungsportfolio von Atos gehört auch die Erstellung von skalierbaren Ausführungsplattformen wie etwa Hadoop. Atos bietet hier verschiedenste Dienstleistungen an.

	Management	Auf der Management-Ebene bietet Atos Outsourcing Services an, welche die verschiedensten Bereiche abdecken. Hierbei werden Rechenzentren zur Verfügung gestellt, welche sowohl eine hohe Speicherkapazität als auch eine hohe Rechenleistung bieten. Je nach Dienstleistung kann hier auch ein auf die Bedürfnisse zugeschnittene Datenbank ausgeliefert werden.

Attachmate Group

Attachmate ist eines von 4 Unternehmen der "Attachmate Group". Zu dieser Gruppe gehören noch folgende Unternehmen: Novell, NetIQ und SUSE. Die Gruppe gehört Privatpersonen und ist nicht am Aktienmarkt vertreten. Attachmate ist in insgesamt 40 Ländern vertreten und gehört zu den größten Softwareunternehmen der Welt. Das Unternehmen besteht bereits seit 1982 und hat seither den Fokus darauf, "die richtigen Informationen zur richtigen Zeit in den richtigen Formaten auf die richtigen Geräte" zu liefern. Attachmate betreibt eine Niederlassung in Wien, welche sich vorrangig um den Verkauf der Produkte und Lösungen kümmert.

Utilization	-
Analytics	-
Platform	Das Produkt „Attachmate Luminet“ hat die Aufgabe, Betrugsdelikte und Datenmissbrauch zu stoppen. Hierbei wird das Verhalten von Benutzern überwacht und analysiert.
Management	-

BMC

BMC Software ist ein Anbieter von Enterprise Management-Lösungen. Die Lösungen bieten vorausschauende Optimierung von IT-Services, verwalten geschäftskritische Anwendungen und Datenbanken, reduzieren Kosten und steigern die Leistung komplexer IT-Architekturen. Das Unternehmen verfügt über etwa 20 Mitarbeiter bei rund 5 Mio Euro Umsatz

Utilization	Atrium IT Dashboards and IT Analytics bietet Dashboards für Entscheidungsträger. Hierbei werden umfangreiche Dashboards für die Entscheidungsfindung angeboten.
Analytics	-
Platform	-
Management	BMC bietet mehrere Produkte im Bereich „Management“ an. Hierzu zählen: • Bladelogic Database Automation. Dieses

Produkt bietet eine Automatisierung für Datenbanken, welches vor allem für skalierbare Systeme von Vorteil ist.

- Bladelogic Server Automation. Dieses Produkt bietet eine Serverautomatisierung an, welche von skalierbaren Plattformen wie etwa Hadoop verwendet werden kann.

Braintribe

Braintribe ist ein österreichischer Anbieter, der ursprünglich aus dem ECM Bereich kommt. Neben dem Headquarter in Wien bestehen auch Niederlassungen in Deutschland, der Schweiz, Brasilien und den Vereinigten Staaten.

Utilization	-
Analytics	-
Platform	Braintribe bietet mit der Plattform „Tribefire“ eine Lösung an, welche für „intelligentes Informationsmanagement“ steht. Der Fokus der Anwendung ist, das immer mehr Anwendungen im Unternehmen existieren, welche Daten in unterschiedlichen Formaten und Quellen abspeichern. Diese Daten müssen schlussendlich kombiniert werden, damit ein Mehrwert generiert werden kann. Das ist die Aufgabe von Tribefire.
Management	-

Catalysts

Catalysts ist ein junges österreichisches Software Entwicklungs-Unternehmen mit Standorten in Wien, Linz, Hagenberg und Rumänien. Das Unternehmen ist ein österreichisches Start-Up mit rund 35 Mitarbeitern.

Utilization	-
Analytics	-
Platform	Die Catalysts GmbH bietet Softwaredienstleistungen an, jedoch keine eigenständigen Produkte. Da das Unternehmen sehr stark im Bereich für verteilte Systeme tätig ist, hat die Firma Catalysts einen eigenen R&D Cluster für Big Data eingerichtet. Das Unternehmen hat bereits mehrere Projekte im Hadoop-Umfeld sowie im High-Performance Processing Umfeld durchgeführt.
Management	-

CommVault

CommVault Systems wurde 1996 gegründet, und hat sein

Headquarter in Oceanport, New Jersey. Das Unternehmen erwirtschaftete zuletzt weltweit einen Umsatz von rund 400 Millionen USD und beschäftigte etwa 1500 Mirarbeiter. In Österreich verfügt das Unternehmen über keine Repräsentanz, ist aber trotzdem am Markt aktiv.

Utilization	-
Analytics	-
Platform	-
Management	CommVault bietet mit der Plattform „Simpana“ eine umfangreiche Lösung an, welche das Management von Dateidaten erleichtert. Hierbei kann man z.B. E-Mails auf einfache Art und Weise archivieren und auch wieder schnell darauf zugreifen.

Compuware

Die Compuware Corporation ist ein US-amerikanisches Softwareunternehmen mit Hauptsitz in Detroit, Michigan. Es wurde 1973 von Peter Karmanos Jr., Thomas Thewes und Allen Cutting gegründet. Compuware beschäftigt weltweit rund 4600 Mitarbeiter (Stand 2012). In Österreich verfügt das Unternehmen über einen Standort in Linz und Wien.

Utilization	-
Analytics	-
Platform	-
Management	Compuware bietet die Anwendung „Compuware APM“ für verschiedene Big Data Plattformen wie etwa Hadoop, Cassandra oder Splunk an. Hierbei handelt es sich um eine Software, welche sich um das Performance-Management der eingangs erwähnten Produkte handelt.

CSC

CSC ist ein globales IT-Beratungs- und Dienstleistungsunternehmen mit fundiertem, branchenspezifischem Know-how, das herstellerunabhängig und end-to-end die für den Kunden beste Lösung bietet. 80 MEU, 350 Mitarbeiter. eutschland, Österreich und Schweiz sind die deutschsprachigen Länder von CSC. Gemeinsam mit Eastern Europe und Italien bilden sie die Region Central & Eastern Europe (CEE).

Utilization	-
Analytics	-
Platform	Das Unternehmen CSC bietet verschiedenste Dienstleistungen rund um das Thema Big Data an, jedoch gibt es keine eigenen Produkte. CSC

	implementiert verschiedene Lösungen auf Basis von Hadoop, welches dem Plattform-Level zugeordnet wird.
Management	CSC bietet verschiedenste Outsourcing Services an, welche für Big Data Lösungen verwendet werden. Hierbei werden komplette Rechenzentren erstellt.

Dassault Systemes

Das Unternehmen "Dassault Systems" wurde 1981 als Spin-Off des Unternehmens "Dassault Aviation" gegründet. Hierbei waren einige Ingenieure des alten Unternehmens an der Gründung beteiligt. Der damalige Fokus des Unternehmens war die 3D-Visualisierung, welches noch heute ein wichtiger Bereich des Unternehmens ist. Der Börsengang des Unternehmens erfolgte im Jahr 1996. Dassault Systems verfügt über eine Niederlassung in Österreich.

Utilization	Das Netvibes Dashboard von Dassault Systemes bietet eine Entscheidungsunterstützung basierend auf Dashboards an.
Analytics	Dassault Systemes bietet auf der Ebene "Analytics" das Produkt "Exalead" an. Hierbei wird aufgrund einer umfassenden Datenbasis neue Information gewonnen.
Platform	-
Management	-

Dell

Dell Inc. (vormals Dell Computer Corporation) ist ein US-amerikanischer Hersteller von Computern. Das Unternehmen hat seinen Hauptsitz in Round Rock, Texas. 1984 gegründet, fast 60 Mrd. USD Umsatz, weltweit 110.000 Mitarbeiter. Die österreichische Niederlassung befindet sich in Wien, rund 40 Mitarbeiter erwirtschaften lokal rund 6 Millionen Euro Umsatz.

Utilization	-
Analytics	-
Platform	Dell bietet auf Basis von Hadoop eine umfassende Lösung an, welche Cloudera verwendet. Hierbei kommt nicht nur Hadoop zum Einsatz sondern auch Speichersysteme von Dell. Damit ist das Produkt, welches Dell hier anbietet über 2 Bereiche verteilt.
Management	Dell bietet für deren Big Data Lösung nicht nur eine Hadoop-Implementierung an, sondern auch die Serversysteme für die Ausführung und Speicherung. Diese basieren auf den Standard-

Serversystemen von Dell.

Dropbox

Dropbox ist ein 2007 gegründeter Webdienst, der die Synchronisation von Dateien zwischen verschiedenen Computern und Personen ermöglicht. Er kann damit auch zur Online-Datensicherung verwendet werden. Der Zugriff auf Dropbox ist im Browser und mit Hilfe von Anwendungen für verschiedene Betriebssysteme möglich. Dropbox hat seinen Sitz in den Vereinigten Staaten und greift selbst auf einen Speicherdienst von Amazon zurück. Das Unternehmen ist nicht in Österreich tätig.

Utilization	-
Analytics	-
Platform	-
Management	Dropbox bietet einen Speicherdienst in der Cloud an. Ursprünglich sehr populär als freie Version verwenden auch immer mehr Unternehmen die Cloud Speicher von Dropbox.

EMC

Die EMC Corporation ist ein US-amerikanischer Hersteller von Hardware und Software mit Unternehmenssitz in Hopkinton, Massachusetts. EMC betreibt Niederlassungen in über 60 Ländern, wie beispielsweise Großbritannien, Deutschland, Frankreich oder Ägypten. EMC ist spezialisiert auf Speichersysteme für Betreiber von großen Rechenzentren im Enterprise-Bereich, Software zum Dokumenten- und Content-Management (insbesondere die Documentum-Produktpalette) und Backup-Lösungen . Weitere wichtige Themen sind Virtualisierung, Informationssicherheit und Big Data. In Österreich hat das Unternehmen mehr als 500 Mitarbeiter bei über 100 Millionen Euro Umsatz.

Utilization	-
Analytics	-
Platform	Mit dem Produkt "Greenplum" bietet EMC eine skalierbare Ausführungsplattform auf Basis von Hadoop an. Hierbei werden sämtliche Bereiche von Hadoop abgedeckt.
Management	EMC bietet eine ganze Reihe an Produkten für die Management-Ebene an. Hierzu zählen: • EMC Isilon OneFS Operating System. Dieses Produkt erstellt ein eigenständiges Dateisystem über verschiedene Systeme in einem Cluster und macht es so einfacher ansprechbar. • EMC Isilon Platform Nodes and Accelerators. Dieses Produkt liefert eine

Speicherlösung für ein einzelnes System und ist primär auf I/O intensive Operationen ausgelegt.
- EMC Isilon Infrastructure Software. Diese Software hat die Aufgabe, Daten effektiv und kostengünstig zu sichern.

Fabasoft Group

Fabasoft ist ein Softwarehersteller mit Sitz in Linz, Oberösterreich. Das Unternehmen wurde 1988 von Helmut Fallmann und Leopold Bauernfeind gegründet. Der Name Fabasoft ist eine Abkürzung aus Fallmann Bauernfeind Software. Die Aktien der Fabasoft AG notieren seit dem 4. Oktober 1999 im Prime Standard der Frankfurter Wertpapierbörse. Das Unternehmen kann in Österreich auf einen Umsatz von 23 Millionen Euro verweisen, der mit rund 200 Mitarbeitern erwirtschaftet wird.

Utilization	-
Analytics	-
Platform	Fabasoft selbst bietet keine Big Data Lösungen an, jedoch hat das Unternehmen ein Tochterunternehmen namens "Mindbreeze". Die Lösung des Unternehmen ist weiter unten beschrieben.
Management	-

Fujitsu

Fujitsu ist ein im Nikkei 225 gelisteter japanischer Technologiekonzern. Schwerpunkte der Produkte und Dienstleistungen sind Informationstechnologie, Telekommunikation, Halbleiter und Netzwerke. In Deutschland hat Fujitsu Standorte und Produktionsstätten in Augsburg, Berlin, Dresden, Düsseldorf, Frankfurt, Hamburg, Laatzen, Mannheim, München, Nürnberg, Paderborn, Sömmerda, Stuttgart und Walldorf. Die österreichische Niederlassung verfügt über ca 250 Mitarbeiter bei einem Umsatz von über 140 Millionen Euro.

Utilization	-
Analytics	-
Platform	Mit Interstage Big Data Parallel Processing Server bietet Fujitsu eine Hadoop-Platform an. Hierbei erweitert Fujitsu die Plattform um einen verteilten Datenspeicher, um Daten zwar auf Einzelsystemen zu halten, diese aber öffentlich zugänglich zu machen.
Management	Interstage Terracotta BigMemory Max ist ein verteilter In-Memory Cache, welcher schnellen Zugriff auf Daten ermöglicht. Hierbei können

mehrere Terabyte an Daten abgelegt werden.

Google Inc.

Google Inc. ist ein Unternehmen mit Hauptsitz in Mountain View (Kalifornien, USA), das durch Internetdienstleistungen – insbesondere durch die gleichnamige Suchmaschine „Google“ – bekannt wurde. Gegründet wurde das Unternehmen am 4. September 1998, der jährliche Umsatz betrug zuletzt etwa 50 Mrd. USD, bei mehr als 50000 Mitarbeitern. Das Unternehmen ist auch in Österrreich ansässig, - die Google Austria GesmbH hält einen Standort in Wien.

Utilization	-
Analytics	-
Platform	Google BigQuery liefert einen einfachen Abfragemechanismus für große Datenmengen. Das Produkt setzt hierfür das MapReduce Verfahren ein, welches Google auch für die Suche verwendet. Google BigQuery ist Bestandteil der Google Cloud Produkte. BigQuery liefert vor allem eine skalierbare Ausführungsplattform.
Management	Ein Teilprodukt der Google AppEngine ist der Datastore sowie die Datenbank, welche vor allem auf rießige Datenmengen ausgelegt sind. Hierbei kommt im Falle der Datenbank eine NoSQL-Datenbank zum Einsatz.

HDS

Im Jahre 1959 wurde die Hitachi America Ltd. gegründet, die europäische Niederlassung Hitachi Europe im Jahr 1982. Heute zählt die Hitachi Ltd. Corporation als großer Mischkonzern zu den 100 größten Unternehmen der Welt. In Österreich hat das Unternehmen mehr als 100 Mirarbeiter bei einem Umsatz von rund 50 Millionen Euro. Niederlassungen bestehen in Wien und Linz.

Utilization	-
Analytics	-
Platform	-
Management	Die Hitachi virtual Storage Platform ist eine skalierbare Speicherlösung. Hierbei handelt es sich um Hardware, welche Daten über mehrere Systeme verteilen kann.

HP

Die Hewlett-Packard Company,(HP) ist eine der größten US-amerikanischen Technologiefirmen, registriert in Wilmington, Delaware und mit Firmenzentrale in Palo Alto, Kalifornien. HP ist eines der umsatzstärksten IT-Unternehmen der Welt und war das

erste Technologieunternehmen im Silicon Valley. Zu Hewlett-Packard gehören der Computerhersteller Compaq und Palm, ein Hersteller von PDAs und Smartphones. Die deutsche Hauptniederlassung (Hewlett-Packard GmbH) befindet sich in Böblingen, die schweizerische in Dübendorf (Hewlett-Packard (Schweiz) GmbH) und die österreichische in Wien (Hewlett-Packard Gesellschaft m.b.H.) Neben dem österreichischen Firmensitz in Wien bestehen Niederlassungen in Graz, Götzis und St. Florian. In Österreich hat HP an diesen Standorten mehr als 800 Mitarbeiter, und einem Umsatz jenseits der 200 Millionen Euro.

Utilization	-
Analytics	-
Platform	HP bietet umfangreiche Hadoop-Lösungen an, welche auf den Serversystemen von HP zum Einsatz kommen. Hierfür bietet HP einige Referenzarchitekturen an, welche auf die eigenen Systeme ausgelegt sind.
Management	HP bietet eine ganze Reihe an Infrastrukturlösungen für das Datenmanagement an. Hierbei sind vor allem die HP Serversysteme von Interesse. HP bietet auch Dienstleistungen rund um Big Data an, womit ein Bau von Rechenzentren möglich ist.

IBM

Die International Business Machines Corporation (IBM) ist ein US-amerikanisches IT- und Beratungsunternehmen mit Sitz in Armonk bei North Castle im US-Bundesstaat New York. IBM ist eines der weltweit führenden Unternehmen für Hardware, Software und Dienstleistungen im IT-Bereich sowie eines der größten Beratungsunternehmen. Aktuell beschäftigt IBM weltweit 426.751 Mitarbeiter bei einem Umsatz von mehr als 100 Mrd. USD. In Österreich beschäftgt das Unternehmen mehr als 1200 Mitarbeiter bei einem Umsatz von über 400 Millionen Euro. Neben Wien bestehen Geschäftsstellen in Oberösterreich, Salzburt, Tirol, Voralberg, Steiermark und Kärnten.

Utilization	-
Analytics	-
Platform	IBM bietet eine ganze Reihe an Services rund um dem Platform-Bereich für Big Data. Hierbei werden vor allem Hadoop-Implementierungen geboten, wobei auch eigene Lösungen für Streaming-Verfahren eingesetzt werden. Diese Lösungen werden vielfach auf die Serversysteme von IBM zugeschnitten.

	Management	IBM bietet zum Einem ein Outsourcing von Rechenzentren an, erstellt aber auch Rechenzentren für Kunden. Die verschiedensten Serversysteme, welche für die Datenanalyse und -speicherung eingesetzt werden können, sind auch in den Rechenzentren der IBM vorhanden.

Infor

Infor ist ein weltweiter Anbieter von Geschäftssoftware für spezielle Industrien (13 derzeit). Mit einem Umsatz von rd. 2,8 Mrd. US-Dollar ist Infor der inzwischen drittgrößte Hersteller neben SAP und Oracle. Infor hat lt. aktuellen Zahlen vom Juni 2012 weltweit etwa 70.000 Kunden und Niederlassungen in 164 Ländern und arbeitet mit 1500 Partnern zusammen. Der Unternehmenshauptsitz war ursprünglich durch den Gründer und Ex-CEO Jim Schaper in Alpharetta, Georgia, Vereinigte Staaten. n Österreich bestehen kleine Niederlassungen in Wien und Linz.

Utilization	Infor bietet eine breite Palette an Lösungen für den Utilization-Layer. Hierzu zählen folgende Anwendungen: • Infor Reporting stellt ein umfassendes Tool für Berichte dar, welche für die Entscheidungsfindung herangezogen werden. • Infor ION Business Analytics ist ein Tool, welches Rollenbasierte Berichte erstellt und flexibel an verschiedene Branchen anpassbar ist. • Infor ION BI ist ein Informationssystem, welches mit einer In-Memory Datenbank arbeitet. Grundlage hierfür ist die Informationsbereitstellung, welche über verschiedene Geräteformen ermöglicht wird. • Infor ION Business Vault überwacht und steuert Geschäftsdaten des Unternehmens in Echtzeit. Daten können in Echtzeit abgefragt werden, was Einblicke in die aktuelle Geschäftstätigkeit erlaubt.
Analytics	-
Platform	-
Management	-

Kofax

Das Unternehmen Kofax, welches aus den USA stammt und am Nasdaq gelistet ist, bietet vorrangig Lösungen für die Bilderkennung und Prozessautomatisierung an. Das Unternehmen ist in mehr als 30

Ländern vertreten und hat über 1.000 Mitarbeiter. In Österreich besteht auch eine Niederlassung.

Utilization	Die Produktlinie "Altosoft", welche durch eine Firmenübernahme zu Kofax kam, bietet vielfältige BI Möglichkeiten.
Analytics	-
Platform	-
Management	-

Linemetrics

Linemetrics ist ein österreichisches IT Startup aus Hagenberg in Oberösterreich. Gegründet wurde das Unternehmen im Sommer 2012 von Experten aus der industrienahen IT. Diese hatten sich zum Ziel gesetzt, Industrieanlagen durch Sensoren einfacher und besser überwachen zu können. Hierfür wurde ein Produkt entwickelt, welches einfach zu installieren ist und ein Cloud-Interface bietet. Das Unternehmen hat in der etwas mehr als 1-jährigen Firmengeschichte bereits eine Vielzahl an Preisen erhalten.

Utilization	-
Analytics	Die Analyseplattform bietet eine Echtzeitanalyse der jeweiligen Metriken, welche von den Geräten geliefert wurden. Hierbei kann man aufgrund gewisser KPIs auf Ereignisse reagieren und den Prozess verbessern.
Platform	-
Management	Linemetrics bietet eine Hardware, welche einfach an die jeweiligen Geräte und Sensoren anschließen kann. Diese kommuniziert in weitere Folge mit der Analyseplattform.

Microsoft

Die Microsoft Corporation ist ein multinationaler Software- und Hardwarehersteller. Mit 94.290 Mitarbeitern und einem Umsatz von 73,72 Milliarden US-Dollar ist das Unternehmen weltweit der größte Softwarehersteller. Der Hauptsitz liegt in Redmond, einem Vorort von Seattle (US-Bundesstaat Washington). IN Österrech erwirtschaftet das Unternehmen mit rund 350 Mitarbeitern etwa 400 Millionen Umsatz. Seit 1991 ist Microsoft mit einer eigenen Niederlassung in Wien vertreten, seit 2006 verfügt das Unternehmen mit Vexcel Imaging über eine F&E-Niederlassung in Graz.

Utilization	-
Analytics	-
Platform	Auf Basis von Windows Azure bietet Microsoft ein Service an, welches Hadoop verwendet. Hierbei wird die gesamte Umgebung verwaltet.

	Management	Microsoft bietet eine ganze Reihe an Management-Tools an. So ist das System Center ein wichtiges Tool, um eine große Anzahl an Systemen zu steuern und zu überwachen. Microsoft bietet mit dem SQL Server eine Relationale Datenbank an, welche sich auch für verteilte, hochskalierbare Szenarien eignet. Mit Windows Azure, dem Cloud Angebot von Microsoft, bietet Microsoft mit dem Windows Azure Table Storage eine nichtrelationale Datenbank an.
MicroStrategy	Das Unternehmen "MicroStrategy" fokusiert auf Business Intelligence und Big Data. MicroStrategy ist am Nasdaq gelistet und hat weltweit über 3.200 Mitarbeiter in 26 Ländern. In Österreich hat das Unternehmen eine Niederlassung in Wien.	
	Utilization	Das Unternehmen bietet hierbei eine umfassende Lösung Für Business Intelligence an. Der Fokus richtet sich auf das Berichtswesen im Unternehmen wie beispielsweise Scorecards.
	Analytics	Das Unternehmen MicroStrategy bietet eine Lösung für Big Data an. Hierbei werden verschiedene Datenquellen als Grundlage verwendet. Die in-Memory Technologie des Unternehmens sorgt für eine schnelle und übersichtliche Auswertung.
	Platform	-
	Management	-
mindbreeze	Fabasoft Mindbreeze ist eine Software-Produktlinie für Enterprise Search, Information Access und Digital Cognition. Das Flagship-Produkt ist Fabasoft Mindbreeze Enterprise (vormals Mindbreeze Enterprise Search). Entwickelt werden die Produkte von der Mindbreeze GmbH, einem Tochterunternehmen der Fabasoft-Gruppe mit Sitz in Linz, Oberösterreich. Siehe Unternehmensdetails zu Fabasoft.	
	Utilization	-
	Analytics	-
	Platform	Das zur Fabasoft Gruppe gehörende Unternehmen Mindbreeze bietet eine

		umfassende Plattform zur Suche von Informationen. Diese können auf einfache Art und Weise in Websites und Anwendungen integriert werden.
	Management	-
NetApp	NetApp, Inc., zuvor bekannt als Network Appliance, Inc., ist ein Unternehmen, das im Bereich der Datenspeicherung und des Datenmanagements arbeitet. Es hat seinen Sitz in Sunnyvale im US-Bundesstaat Kalifornien. Das Unternehmen wird im NASDAQ-100 geführt. Das Unternehmen verfügt über eine Niederlassung in Wien	
	Utilization	-
	Analytics	-
	Platform	-
	Management	NetApp bietet eine ganze Reihe von skalierbaren Speichersystemen an. Der Fokus hierbei liegt auf große Datenmengen.
OpenLink	OpenLink bietet Produkte im Business Intelligence und Transaction Lifecycle Management bereich an. Das Unternehmen existiert bereits seit mehr als 20 Jahren und hat über 1.200 Mitarbeiter weltweit. OpenLink fokusiert auf die Industriezweige "Energie", "Finanzen" und "Handel". In Österreich hat das Unternehmen eine Niederlassung in Wien	
	Utilization	OpenLink bietet spezielle Lösungen für die jeweiligen Industrien an, in welchen das Unternehmen tätig ist. Hierbei bietet das Unternehmen verschiedene BI Lösungen, welche vor allem auf Risikooptimierung ausgerichtet sind.
	Analytics	-
	Platform	-
	Management	-
OpenText	"Die OpenText Corporation ist das größte kanadische Softwareunternehmen. Die Firma ist an der Toronto Stock Exchange und der NASDAQ gelistet. Die OpenText Corp. wurde 1991 gegründet, erwirtschaftete 2012 Umsätze von rund 1,2 Mrd. USD und beschäftigt etwa 5000 Mitarbeiter (2013). Das Unternehmen verfügt auch über eine Niederlassung in Wien mit rund 30 Mitarbeitern werden rund 15 Millionen Euro erwirtschaftet."	
	Utilization	-

	Analytics	Mit „Auto-Classification“ steht ein Tool zur Verfügung, welches auf Basis von Algorithmen eine Klassifikation vorschlägt. Diese Klassifikation kann die Arbeit von denjenigen erleichtern, die eben diese verwalten müssen. Umfassende Einblicke können mit dem Produkt „Content Analytics“ erstellt werden. Hierbei wird auf Basis des Content Management Tools von OpenText eine Analyse-Plattform angeboten.
	Platform	Die „Information Access Platform“ von OpenText bietet eine umfassende Plattform für die Datenanalyse auf Basis von OpenText. Hierbei kommen nicht nur analytische Insights sondern auch Datenmigrationen und Archivierungen zum Einsatz.
	Management	
Oracle	Oracle ist einer der weltweit größten Softwarehersteller. Seinen Hauptsitz hat das Unternehmen in Redwood Shores (Silicon Valley, Kalifornien). Seit der Übernahme von Sun Microsystems im Januar 2010 hat Oracle das Portfolio um Hardware erweitert. Gegründet wurde das Unternehmen von Lawrence J. Ellison (Larry Ellison), der bis heute den Vorsitz der Firma hat. Oracle beschäftigt mehr als 115.000 Mitarbeiter und hat 390.000 Kunden in 145 Ländern. Die Deutschlandzentrale ist in München, seit 2001 gibt es Oracle Direct in Potsdam. In Deutschland gibt es zehn Geschäftsstellen. In der Schweiz hat das Unternehmen vier, in Österreich eine Niederlassung in Wien."	
	Utilization	Das Unternehmen Oracle bietet eine Lösung für Business Intelligence an. Hier handelt es sich primär um Enterprise Performance Management Tools basierend auf den Datenbanken Oracle.
	Analytics	-
	Platform	-
	Management	Das Unternehmen bietet viele Lösungen für das Management von Big Data Lösungen an. Hierbei handelt es sich zum einen um die Datenbanken des Unternehmens wie etwa der RDBMS Datenbank, aber auch um verschiedene Nicht relationale Datenbanken Wie Oracle NoSQL. Durch die Übernahme von SUN Microsystems hat Oracle das Portfolio auch um

Hardwarelösungen erweitert. Hierbei handelt es sich um Serversysteme welche für Big Data Lösungen eingesetzt werden können.

Parstream

Parstream ist ein erfolgreiches Startup mit Niederlassungen in Silicon Valley, Deutschland und Frankreich. Der österreichische Markt wird von der deutschen Niederlassung aus mitbetreut.

Utilization	-
Analytics	Das Unternehmen Parstream bietet eine Lösung für die Echtzeitanalyse von Daten an. Die Antwortzeiten der Datenbank für Analysen bewegen sich im Millisekunden Bereich.
Platform	-
Management	-

Pitney Bowes Software

Die Kernkompetenz von Pitney Bowes Software ist das Datenmanagement von großen Datenbeständen. Fokusgebiete umfassen hierbei Customer & Marketing Analytics, Datenmanagement und Kampagnenmanagement. Für all diese Bereiche werden große Datenmengen verarbeitet. In Österreich gibt es eine eigene Niederlassung. 2012 betrug der Jahresumsatz 4.9 Milliarden US-Dollar, der operative Gewinn des Unternehmens belief sich auf 769 Millionen $. Das Unternehmen besteht bereits seit 1920 und ist eines der Fortune 500 Unternehmen. Die ursprüngliche Kernkompetenz des Unternehmens stammt aus dem Postsegment. Noch heute bietet das Unternehmen eine große Produktpalette für Frankiermaschinen und Sortiersysteme an.

Utilization

"Pitney Bowes bietet auf der Ebene der Utilization gleich mehrere Produkte an. Hierbei gibt es eine ganze Reihe an Marktanalyseprodukte. Diese sind:

- der Portrait Explorer, welcher Analysefunktionen für Kundendaten im Unternehmen bietet
- der Portrait Miner zur Analyse von Kampagnen
- Portrait Uplift, welches eine bessere Zielgruppenorientierung der Kunden anbietet.

Im Location Intelligence Bereich bietet Pitney Bowes umfangreiche Lösungen für Datenanalysen im Umfeld von Geografiedaten. Hierfür bietet das Unternehmen die Produkte

	"MapInfo Suite" und "Enterprise Location Intelligence" an. Weitere Services von Pitney Bowes umfassen Lösungen für das Datenmanagement. Hierbei gibt es umfangreiche Lösungen für Datenintegration und Datenqualitätsmanagement. Für das Kampagnenmanagement bietet Pitney Bowes folgende Produkte an: • Portrait Dialogue: diese Software steuert und überwacht cross-mediale Kampagnen unter Einbezug von Social Media Kanälen • Portrait Interaction Optimizer: ermöglicht die Ansprache der Kunden über die nächsten besten Möglichkeiten • Portrait Foundation: ist ein Framework, welches einfache Funktionsbausteine für CRM-Anwendungen bietet.
Analytics	-
Platform	-
Management	-

QlikTech

QlikTech wurde 1993 in Schweden gegründet, hat jedoch mittlerweile den Hauptsitz in den USA. Derzeit beschäftigt das Unternehmen über 1.600 Mitarbeiter weltweit und bedient damit mehr als 30.000 Kunden. Das Credo von QlikTech lautet "Simplifying decisions for everyone, everywhere". QlikTech hat eine österreichische Niederlassung in Wien.

Utilization	Die Lösung des Unternehmens fokussiert auf die Präsentation von Business Intelligence Daten. Diese werden auf vielen gängigen Geräte Klassen dargestellt und bieten somit eine einfache Entscheidungsfindung auf Basis von Daten.
Analytics	-
Platform	-
Management	-

SAP

Die SAP Aktiengesellschaft mit Sitz im baden-württembergischen Walldorf ist der nach Umsatz größte europäische (und außeramerikanische) Softwarehersteller. Tätigkeitsschwerpunkt ist die Entwicklung von Software zur Abwicklung sämtlicher Geschäftsprozesse eines Unternehmens wie Buchführung,

Controlling, Vertrieb, Einkauf, Produktion, Lagerhaltung und Personalwesen. Das Unternehmen hat weltweit mehr als 60000 Mitarbeiter bei einem Umsatz von 16Mrd. EUR. In Österreich werden rund 400 Mitarbeiter bei einem Umsatz von rund 180 Millionen Euro beschäftigt. Sitz der Österreich-Niederlassung ist Wien.

Utilization	-
Analytics	SAP bietet eine ganze Reihe an Lösungen für Analytics an. Diese sind oftmals auf den Systemen von SAP aufgebaut. Zum Einsatz kommt hierbei Data Warehouse sowie Daten aus den ERP Systemen von SAP. Das Unternehmen bietet auch Lösungen für Predictive Analytics an.
Platform	-
Management	Kernprodukt von SAP ist die Datenbank HANA. Diese Datenbank kombiniert verschiedenste Techniken um zum einem eine hohe Performance durch einen Columnstore bereitzustellen, aber andererseits auch eine in-Memory Technologie für Business Analytics bereitzustellen.

SAS

SAS Institute ist ein 1976 gegründetes, weltweit operierendes Softwarehaus in Privatbesitz mit Sitz in Cary, North Carolina, USA mit 2,87 Mrd US$ Jahresumsatz im Jahr 2012. Damit ist das SAS Institute eines der größten Softwareunternehmen in Privatbesitz und einer der größten Business-Analytics-Anbieter weltweit. Die deutsche Hauptniederlassung befindet sich in Heidelberg, das Hauptbüro für Österreich in Wien, und das Hauptbüro für die Schweiz in Wallisellen in der Nähe von Zürich. Die österreichische Niederlassung umfasst rund 40 Mitarbeiter und ist im wesentlichen eine Vertriebsniederlassung.

Utilization	-
Analytics	SAS bietet primär Lösungen für Business Analytics an. Diese basieren stark auf statistischen Modellen und sind für die verschiedensten Branchen gedacht. Kernprodukt von SAS ist SAS Analytics.
Platform	-
Management	-

simplectix

Simplectix ist ein Startup aus Linz in Oberösterreich. Das Unternehmen fokusiert auf Big Data und Analytics. Das Unternehmen will die "Business Intelligence" für die Zukunft zur Verfügung stellen. Gegründet wurde das Unternehmen im Jahr 2012.

Utilization	-
Analytics	Das Unternehmen bietet eine Softwarelösung Für Predictive Analytics und Mining an. Das Produkt bietet Einblick in verschiedenste Muster. Ziel des Produktes ist es die eigenen Kunden und den Markt besser zu verstehen.
Platform	-
Management	-

Software AG

Die Software AG mit Sitz in Darmstadt ist ein Anbieter für Softwarelösungen verbundene Dienstleistungen für Unternehmen . Ihre Produkte ermöglichen es, Geschäftsprozesse zu analysieren und zu verwalten und IT-Infrastrukturen zu steuern, wobei offene Standards verwendet werden. Das Unternehmen ist das zweitgrößte Softwarehaus in Deutschland und das viertgrößte in Europa.
Per Ende Dezember 2012 erzielte das Unternehmen mit insgesamt 5419 Mitarbeitern einen Konzernumsatz in Höhe von 1,047 Milliarden Euro. In Österreich erwirtschaftete das Unternehmen zuletzt über 3o Mio Euro Umsatz mit rund 100 Mitarbeitern. Die österreichische Niederlassung hat ihren Sitz in Wien.

Utilization	-
Analytics	Das Produkt Apama Real Time Analytics bietet eine Echtzeitanalyse von Daten. Damit werden Einblicke in verschiedenste Metriken geliefert. Das Produkt macht es möglich dass man schnell auf wichtige Ereignisse reagieren kann und somit wichtige Geschäftsentscheidungen treffen kann.
Platform	-
Management	Mit dem Produkt Terracotta bietet die Software AG eine in-memory Datenbank Lösung an. Diese Datenbank kann vor allem für große Datenmengen eingesetzt werden.

Symantec

Die Symantec Corporation ist ein US-amerikanisches Softwarehaus, das im Jahr 1982 gegründet wurde. Es ist seit dem 23. Juni 1989 an der NASDAQ börsennotiert. Der Hauptsitz des Unternehmens liegt seit dem 5. Oktober 2009 in Mountain View (Kalifornien/USA), in der Nähe des geografischen Zentrums des Silicon Valley. Symantec betreibt nach eigenen Angaben Niederlassungen in 40 Ländern und

beschäftigt insgesamt etwa 18.500 Mitarbeiter. Das Österreich-Business wird von der Niederlassung in Wien organisiert.

Utilization	-
Analytics	-
Platform	-
Management	Symantec bietet eine Archivierungsplattform für Big Data Lösungen. Hierbei sind vor allem die Archivierung von verteilten Daten eine wichtige Komponente. Das Produkt von Symantec nennt sich NetBackup.

Syncsort

Syncsort hat sein Headoffice in Woodcliff Lake, New Jersey, United States. Darüber hinaus bestehen Niederlassungen in Frankreich, Deutschland und UK, mit einem internationalen support center in Holland. Syncsort Produkte werden über Distributoren und Parnter in Österreich vertrieben, das Unternehmen besitzt keine österr. Niederlassung.

Utilization	-
Analytics	-
Platform	SyncSort bietet mit dem Produkt DMX-h eine Lösung für Hadoop an. Diese setzt vor allen Dingen auf eine einfachere Verwendung von Hadoop Lösungen. Mit diesem Produkt ist es möglich dass man keine Zeile Code schreiben muss. Man bedient sich lediglich eines visuellen Designers.
Management	-

Teradata

Teradata hat sich zum 1. Oktober 2007 von der NCR Corporation abgespaltet und ist nun ein eigenständiges, börsennotiertes Unternehmen. Das Unternehmen ist weltweit in zahlreichen Ländern vertreten, erwirtschaft einen Umsatz von weit über Mrd. USD und hat auch eine Niederlassung in Wien.

Utilization	Das Produkt Teradata Decision Experts bietet eine Unterstützung für die Erstellung von Berichten für die Finanzdaten eines Unternehmens. Dadurch können Entscheidungen besser getroffen werden.
Analytics	Teradata hat eine ganze Menge an Analytics-Anwendungen im Portfolio. Hierzu zählen: • Teradata Master Data Management • Teradata Demand Chain Management • Teradata Value Analyzer

- Teradata Warehouse Miner

Platform Teradata bietet eine Vielzahl an Lösungen an, welche auf Hadoop basieren. Hierzu zählen:

- Teradata Appliance for Hadoop
- Teradata Aster Big Analytics Appliance
- Teradata Enterprise Access for Hadoop

Management Das ursprünglich Produkt des Unternehmens ist eine Datenbank, welche auch heute noch verfügbar ist. Diese gliedert sich in folgende Datenbanken:

- Teradata Database. Die ursprüngliche Datenbank in der aktuellen Version.
- Teradata Columnar. Eine hoch performante Datenbank welche nicht auf Reihen sondern auf Spalten selektiert. Danach ist die Geschwindigkeit bei Abfragen sehr hoch. Im Grunde genommen handelt es sich um einen Key/Value Speicher.
- Teradata Temporal. Eine temporäre Datenbank welche die Zeitdimension berücksichtigt. Dadurch ist es möglich Veränderungen von Datensätzen über einen Zeitraum darzustellen und zu analysieren.
- Teradata Intelligent Memory. Diese Datenbank speichert die wichtigsten Informationen im Hauptspeicher wodurch der Zugriff auf diese beschleunigt wird.

Veeam

Veeam Software ist ein Schweizer Hersteller von Software mit Hauptsitz in Baar ZG. Der Vertrieb ist dezentral organisiert und erfolgt über verschiedene Partner. Diese vertreiben sowohl die Softwarelizenzen wie auch den Veeam Service. Der nord-amerikanische Hauptsitz befindet sich in Columbus, Ohio USA. Stammsitz für den asiatisch-pazifischen Raum ist Sydney, Australien. Veeam hat weltweit mehr als 57.000 Kunden. In Österrreich besteht keine Niederlassung.

Utilization -

Analytics -

Platform -

Management Das Unternehmen bietet eine Speicher- und Backuplösung Für Unternehmensdaten in virtuellen und verteilten Systemen an. Des weiteren gibt es eine umfangreiche Lösung für virtuelle Maschinen, Welche auch gespeichert werden können.

visalyze

Visalyze ist ein Unternehmen aus Innsbruck in Tirol. Das Unternehmen hat es sich zur Aufgabe gemacht, soziale Medien durch Datenanalyse und Datenvisualisierung besser verständlich zu machen. Seit dem Bestehen des Unternehmens hat es bereits eine Vielzahl an Förderungen aquirieren können.

Utilization	Das Produkt des Unternehmens bietet eine umfangreiche Lösung für die Visualisierung von Daten aus sozialen Medien. Hierfür werden Facebook Daten in Echtzeit analysiert und präsentiert. Das bietet einen umfangreiches Einblick in die Unternehmenswahrnehmung und Produktwahrnehmung für potentielle Kunden.
Analytics	-
Platform	-
Management	-

VMware

Vmware ist ein US-amerikanisches Unternehmen, das Software im Bereich der Virtualisierung entwickelt. Das Unternehmen wurde 1998 mit dem Ziel gegründet, eine Technik zu entwickeln, virtuelle Maschinen auf Standard-Computern zur Anwendung zu bringen. VMware ist mit einer Niederlassung in Wien in Österreich vertreten.

Utilization	-
Analytics	-
Platform	-
Management	Das Kernprodukt des Unternehmens ist Virtualisierung, welches auch für Big Data Anwendungen von Interesse ist. Das Unternehmen bietet nicht nur Betriebssystemvirtualisierung sondern auch Speichervirtualisierung an.

webLyzard

Das Unternehmen webLyzard stammt aus Wien in Österreich und liefert Big Data Produkte. Das Unternehmen wurde von Arno Scharl, einem Professor an der Modul Universität Wien gegründet. webLyzard hat bereits eine Vielzahl von Förderungen für deren Produkt bekommen.

Utilization	-
Analytics	Das Produkt beschäftigt sich hauptsächlich mit Nachrichtenanalyse und der Analyse von sozialen Medien, damit man besseren Einblick in aktuelle Geschehnisse erhalten kann. Weitere Bereiche des Produktes sind das Wissensmanagement und eine intelligente Suche.
Platform	-
Management	-

Tabelle 3: Big Data Anbieter

2.4 Weitere Marktteilnehmer

In dieser Sektion werden weitere Unternehmen aufgeführt, welche in Österreich nicht aktiv sind, jedoch aufgrund verschiedener Tatsachen für den Österreichischen Markt interessant erscheinen. Oftmals handelt es sich hierbei um beliebte Open Source Tools von oftmals international sehr großen und bekannten Unternehmen im Big Data Bereich.

10gen

10gen ist das Unternehmen, welches Hinter MongoDB steht. 10gen hat in Europa 3 Niederlassungen, wobei jedoch keine im deutschsprachigen Raum liegt. MongoDB ist eine führende NoSQL-Datenbank, welche vor allem zur Speicherung von großen Datenmengen gedacht ist. Aufgrund der bedeutung von MongoDB in der NoSQL- und Datenbankwelt wird diese Lösung hier aufgeführt

Utilization	-
Analytics	-
Platform	-
Management	10gen entwickelt die NoSQL Datenbank "MongoDB"

Actuate

Actuate bietet das Produkt "Birt" an, welches vor allem für Unternehmen im Retail-Bereich gedacht ist. Actuate stammt aus Kalifornien und hat eine Niederlassung in Deutschland. Das Unternehmen tritt derzeit nicht aktiv am österreichischen Markt auf.

Utilization	Das Produkt "Birt" liefert eine einfache Plattform zum Designen von Anwendungen. Hierbei steht die Visualisierung und damit die Entscheidungsunterstützung im Vordergrund.
Analytics	-
Platform	-
Management	-

Amazon.com Inc.

Amazon Web Services ist der führende Cloud Computing Anbieter weltweit. Amazon Web Services ist in Deutschland sehr aktiv, wobei

sich die Aktivitäten in Grenzen halten. Aktuell gibt es keine Niederlassung in Österreich, es wird auch kein aktives Marketing oder Verkauf betrieben. Obwohl Amazon Web Services keine Niederlassung in Österreich hat, verfügt das Unternehmen über breite Bekanntheit und einigen Partnern in Österreich. Partner des Unternehmens wickeln einzelne Projekte mit Amazon Web Services ab. Es ist auch aus Österreich möglich, Services von Amazon Web Services zu verwenden.

Utilization	-
Analytics	-
Platform	Amazon bietet auf der Plattform-Ebene skalierbare Ausführungssysteme (Amazon EC2) sowie eine komplette Hadoop-Umgebung mit Amazon Elastic MapReduce an
Management	Amazon bietet auf dieser Ebene verschiedene Speicherlösungen an - diese sind zum einem für Binärobjekte (Amazon S3) und für Datenbanken. Amazon DynamoDB ist eine skalierbare, hochverfügbare Datenbank im NoSQL-Bereich. Mit Amazon Relational Database Service (RDS) bietet Amazon auch eine klassische Datenbank an.

Arcplan

Arcplan bietet BI und Analytic Tools für Unternehmen. Das Unternehmen wurde 1993 gegründet und verfügt über weltweite Niederlassungen. In Österreich gibt es keine eigene Niederlassung.

Utilization	Das Produkt "Arcplan Enterprise" verwendet unterschiedliche Datenquellen, um Analysen in Unternehmen zu ermöglichen. Diese werden primär für die Entscheidungsunterstützung herangezogen.
Analytics	-
Platform	-
Management	-

Actuate

Actuate bietet das Produkt "Birt" an, welches vor allem für Unternehmen im Retail-Bereich gedacht ist. Actuate stammt aus Kalifornien und hat eine Niederlassung in Deutschland. Das Unternehmen tritt derzeit nicht aktiv am österreichischen Markt auf.

Utilization	Das Produkt "Birt" liefert eine einfache Plattform zum Designen von Anwendungen. Hierbei steht die Visualisierung und damit die Entscheidungsunterstützung im Vordergrund.
Analytics	-
Platform	-

	Management	-
Informatica	Informatica ist ein am US-Börsenindex Nasdaq gelistetes Unternehmen. Das Unternehmen beschäftigt aktuell (2013) rund 2.700 Mitarbeiter und hat mehrere Standorte weltweit. Einen Standort in Österreich gibt es aktuell nicht. Informatica hat in der Schweiz und in Deutschland Standorte. Der Fokus des Unternehmens liegt auf Datenintegration	
	Utilization	-
	Analytics	-
	Platform	B2B Data Exchange stellt ein flexibles und skalierbares Datenaustauschsystem für Geschäftspartner dar. Ein Fokus liegt ebenfalls auf Datenereignissen wie z.B. sich ändernde Datenzusammenhänge
	Management	Informatica bietet eine Vielzahl an Produkten für das Datenmanagment und dessen Austausch an. Hierfür steht das Produkt "Application ILM" zur Verfügung, welches für die Lebenszyklusverwaltung von Daten zum Einsatz kommt. Es gibt mehrere Produkte, welche für die Datenqualität bestimmt sind. Ein wichtiges Augenmerk besteht auch für die Datenvirtualisierung
Information Builders Inc.	Information Builders ist ein Unternehmen aus New York, welches im Jahr 1975 gegründet wurde. Der Fokus des Unternehmens liegt auf Business Intelligence, Business Analytics und Datenintegration. Das Unternehmen "Information Builders" hat keine Niederlassung in Österreich, wird jedoch durch Raiffeisen Informatik Consulting repräsentiert.	
	Utilization	Information Builders setzt hauptsächlich auf die Utilization-Schiene. Hierbei bietet das Unternehmen verschiedene Produkte für Business Intelligence und Business Analytics an. Das Kernprodukt hierbei ist "WebFOCUS"
	Analytics	-
	Platform	-
	Management	-
Pentaho	Das Unternehmen Pentaho kommt aus dem US-Bundesstaat Florida. Der Fokus des Unternehmens liegt auf BI-Lösungen und Big Data Analytics. In Österreich gibt es keine aktive Niederlassung, der Markt wird jedoch von München aus adressiert.	
	Utilization	Pentaho bietet ein Produkt für die Business

	Analytics an. Dieses Produkt bietet einen visuellen Designer, welcher für Dahsboards verwendet wird. Das Ziel hierbei ist die Entscheidungsunterstützung zu verbessern.
Analytics	-
Platform	-
Management	-
Revolution Analytics	Revolution Analytics ist das Unternehmen, welches hinter der populären Sprache "R" steht. "R" wird primär für Predictive Analytics und für Statistiken eingesetzt und ist ein Open Source Projekt. Revolution Analytics ist in Österreich nicht aktiv vertreten, aufgrund der Popularität der Programmiersprache "R" findet das Unternehmen hier jedoch Erwähnung. Das Unternehmen ist im Sillicon Valley beheimatet mit Niederlassungen in London und Singapur.
Utilization	-
Analytics	Die Sprache "R" wird als High-Level Query Language kategorisiert. Hierbei handelt es sich um eine Abfragesprache, welche für Statistiken auf große Datenmengen angewendet werden kann.
Platform	-
Management	-
Splunk	Das Unternehmen Splunk bietet die gleichnamige Software zur Analyse von Computergenerierten Daten an. Splunk ist in den USA beheimatet und hat derzeit keine Niederlassung in Österreich. Es befindet sich jedoch eine Niederlassung in Deutschland.
Utilization	Das Produkt "Hunk" setzt auf Hadoop auf und liefert vor allem Dashboards, Datenanalysen und Entscheidungsunterstützungen. Hunk soll die Analyse großer Datenmengen vor allem in einer hinsicht vereinfachen - nämlich jener der Einarbeitungszeit.
Analytics	Das Produkt "Splunk" fokusiert auf die Speicherung, Indizierung und Analyse von Daten aus Computersystemen. Anhand von Mustern können so beispielsweise Probleme in Systemen erkannt werden. Der Vorteil der Software liegt laut Angaben der Hersteller in der schnellen Analyse, womit man keine Tage mehr benötigt um sicherheitsrelevante Informationen zu erhalten.
Platform	-
Management	-

Tabelle 4: Marktteilnehmer

2.4.1 Universitäre und außeruniversitäre Forschung

Die Untersuchungen in dieser Studie zeigen, dass Big Data bei der Österreichischen und auch internationalen Forschungslandschaft ein sehr wichtiges Thema ist. Im Folgenden werden die Ergebnisse der Erhebung der Forschungsinitiativen im Bereich Big Data an Österreichischen Forschungsinstitutionen beschrieben. Dafür wurde Institutionen analysiert und jene identifiziert die Forschung in einer der vier Ebenen des Big Data Stacks (Utilization, Analytics, Platform, Management) durchführen. Da auch die vier Ebenen des Stacks jeweils ein sehr breites Spektrum abdecken, sind zahlreiche Institutionen nur in Teilbereiche dessen tätig. Zusammengefasst wurde festgestellt, dass in der Ebene Utilization vermehrt in den Bereichen Semantic Web und Visualization geforscht wird. Im Bereich Analytics wird auf die detaillierte Erhebung in der Studie (Bierig, et al., 2013) verwiesen. Die Forschungsbeteiligung an Platform und Infrastrukturthematiken ist deutlich geringer. Einige Institutionen decken alle vier Ebenen des Big Data Stacks ab. Darunter befinden sich die Universität Wien, die Technische Universität Wien und die Universität Linz.

	Utilization	Analytics	Platform	Management
Universität Wien				
TU Wien				
WU Wien				
MODUL Universität				
BOKU Wien				
Medizinische Universität Wien				
Max F. Perutz Labs				
Forschungszentrum Telekommunikation Wien (FTW)				
Zentrum für Virtual Reality und Visualisierung (VRVIS)				
Austrian Institute of Technology (AIT)				
Austrian Institute for Artificial Intelligence (OFAI)				
Österreichische Akademie der Wissenschaften				
TU Graz				
Universität Graz				
Medizinische Universität Graz				
Montanuniversität Leoben				
Fraunhofer Austria				

Joanneum DIGITAL				
Know-Center Graz				
Universität Innsbruck				
Medizinische Universität Innsbruck				
UMIT				
Universität Salzburg				
Salzburg Research				
Universität Linz				
Software Competence Center Hagenberg				
IST Austria				
Universität Klagenfurt				
SUMME	**15**	**26**	**5**	**5**

2.4.2 Tertiäre Bildung

Mit über 20 Fachhochschulen und den österreichischen Universitäten gibt es in Österreich ein breites Angebot an tertiärer Bildung. In Bezug auf eine zukunftsorientierte Big Data-relevante Ausbildung wurde in den letzten Jahren der Begriff „Data Scientist“ geprägt. Data Scientist wird als einer der zukunftsträchtigsten Jobs des 21. Jahrhunderts gesehen (Davenport & Patil, 2013) und es wird hier von einer steigenden Nachfrage nach qualifiziertem Personal, das Kompetenzen aus unterschiedlichen Bereichen mitbringt, ausgegangen. Ein Data Scientist muss mit Datenspeicherung, Integration, Analyse und Visualisierung umgehen können und benötigt dafür interdisziplinäre Fähigkeiten aus der Statistik und der Informatik. Generell sollte ein Data Scientist mit allen vier Ebenen des Big Data Stack umgehen können. Eine detaillierte Analyse der benötigten Kompetenzen in Bezug auf das Berufsbild Data Scientists wird in Kapitel 4.2.3 bereitgestellt.

Derzeit werden von mehreren Fachhochschulen und Universitäten in den angebotenen Masterstudiengängen einige der relevanten Themenbereiche abgedeckt. Eine komplette Umsetzung einer Ausbildungsschiene zum Data Scientist konnte von den Studienautoren in Österreich nicht gefunden werden. Die Umsetzung einer kompetenten Ausbildung auf tertiärem Bildungsniveau zum Data Scientist ist aus Sicht des Bereichs Big Data für die Weiterentwicklung des Standorts Österreich essenziell.

Ein Überblick über aktuell angebotene Masterstudiengänge, die Teilbereiche der Big Data-Thematik umfassen, wird in Tabelle 5 dargestellt. Diese Informationen wurden mit Hilfe von Onlinerecherche erhoben. Eine hohe Anzahl an Fachhochschulen und Universitäten deckt Teile des Gebiets ab. Generell können drei Trends erkannt werden: Erstens werden Studien angeboten, welche eine profunde Ausbildung in den Bereichen Management, Platform bis zu Analytics bieten. Diese Studienrichtungen können als „Scientific Computing“ zusammengefasst werden und stellen die effiziente Umsetzung von Technologien und deren Anwendung in den Vordergrund. Zweitens werden einige Studien mit dem Schwerpunkt Information Management angeboten. Diese Studienrichtungen fokussieren auf den Utilization Layer des Big Data Stacks. Drittens werden in der

Tabelle auch Studien aufgelistet, welche einen Fokus auf Big Data-relevante Themenbereiche in der Ausbildung legen.

Tertiäre Bildungseinrichtung	Studiengang	Big Data-relevante Schwerpunkte
Universität Wien	***Scientific Computing***	Scientific Data Management, HPC, Algorithmik
	Computational Science	MathematikerInnen InformatikerInnen NaturwissenschaftlerInnen
TU Wien	***Informatik***	Visual Computing, Computational Intelligence, Software Engineering & Internet Computing
TU Graz	***Informatik***	Knowledge Technologies
JKU Linz	***Computer Science***	Computational Engineering
Technikum Wien	***Softwareentwicklung***	Big Data und Linked Data
FH Wiener Neustadt	***Computational Engineering***	Numerische Modellierung, Mechatronik
FH Joanneum	***Informationsmanagement***	Internetdatenbanken, Data Mining
FH Hagenberg	***Information Engineering und Management***	Wissensmodellierung, Data Engineering
	Software Engineering	Business Intelligence, Machine Learning
FH Vorarlberg	***Informatik***	Big Data Management, Computational Intelligence

Tabelle 5: Übersicht tertiärer Bildungseinrichtungen

2.5 Analyse der vorhandenen öffentlichen Datenquellen

Das Thema Open Data hat in den letzten Jahren in Österreich immer mehr Aufmerksamkeit erregt. Ein wichtiger Ansatz hierbei ist die öffentliche und freie Bereitstellung von Daten durch Behörden der öffentlichen Hand. Dies wird unter dem Begriff Open Government Data zusammengefasst. In einem White Paper zu diesem Thema (Eibl, Höchtl, Lutz, Parycek, Pawel, & Pirker, 2012) werden die Grundsätze von Open Government folgendermaßen dargestellt:

- *Transparenz: stärkt das Pflichtbewusstsein und liefert den Bürgerinnen und Bürgern Informationen darüber, was ihre Regierung und ihre Verwaltung derzeit machen. Die freie Verfügbarkeit von Daten ist eine wesentliche Grundlage für Transparenz.*

- *Partizipation: verstärkt die Effektivität von Regierung und Verwaltung und verbessert die Qualität ihrer Entscheidungen, indem das weit verstreute Wissen der Gesellschaft in die Entscheidungsfindung mit eingebunden wird.*
- *Kollaboration: bietet innovative Werkzeuge, Methoden und Systeme, um die Zusammenarbeit über alle Verwaltungsebenen hinweg und mit dem privaten Sektor zu forcieren.*

Um Open Government-Strategien weiterzuentwickeln, wurde 2013 eine Studie über die wirtschaftlichen und politischen Dimensionen von Open Government Data (OGD) durchgeführt (Hüber, Kurnikowski, Müller, & Pozar, 2013). Diese Studie führt Open Government Data auf die Open Government-Initiative von US-Präsident Barack Obama im Jahre 2009 zurück. Des Weiteren wird OGD als „jene nicht personenbezogenen Daten verstanden, die von der Verwaltung zur freien Nutzung zur Verfügung gestellt werden“ (Dax & Ledinger, 2012). Für die Bereitstellung von Open Data wurde von der Arbeitsgruppe Metadaten eine Metadatenstruktur entwickelt welche in (Cooperation OGD Österreich: Arbeitsgruppe Metadaten, 2013) zur Verfügung steht. Österreich nimmt in diesem Bereich eine führende Position ein und hat hier auch den United Nations Public Service Award 2014 (Bundeskanzleramt, 2014) gewonnen.

2.5.1 Überblick über verfügbare Datenquellen nach Organisation

In Österreich wurde die Open Government Data Plattform data.gv.at gegründet und es wurden Richtlinien für die Bereitstellung von öffentlichen Datenquellen erstellt. Nach diesem White Paper sollen alle zur Verfügung gestellten Datenquellen die folgenden Prinzipien erfüllen (Eibl, Höchtl, Lutz, Parycek, Pawel, & Pirker, 2012): *Vollständigkeit, Primärquelle, zeitnahe Zurverfügungstellung, leichter Zugang, Maschinenlesbarkeit, Diskriminierungsfreiheit, Verwendung offener Standards, Lizenzierung, Dokumentation (Dauerhaftigkeit), Nutzungskosten*. Weitere wichtige Punkte hierbei sind die Verwendung von einheitlichen Bezeichnungen und technische sowie organisatorische Anforderungen. Ein zusätzlicher und wesentlicher Aspekt in Bezug auf Open Data, welcher im Rahmen der Studie häufig genannt wurde, ist Data Governance. Um eine sinnvolle Weiterverwendung zu gewährleisten müssen Daten auch aktuell gehalten werden und in geeigneter Form, zum Beispiel bereinigt, zur Verfügung gestellt werden.

In diesem Rahmen werden in Österreich zurzeit 1119 unterschiedliche Datenquellen (Stand Dezember 2013 – data.gv.at) öffentlich und frei zur Verfügung gestellt und die Tendenz ist stark steigend. Derzeit ist die Menge an zur Verfügung gestellten Daten zwischen den Bundesländern und Städten in Österreich sehr unterschiedlich. Die meisten Datenquellen werden von Wien bereitgestellt, gefolgt von Linz, der Gemeinde Engerwitzdorf, Graz, Oberösterreich und Innsbruck (in dieser Reihenfolge). Ein Überblick über die Anzahl der bereitgestellten Daten kann in Abbildung 7 gefunden werden (Basis sind verfügbare Informationen von data.gov.at).

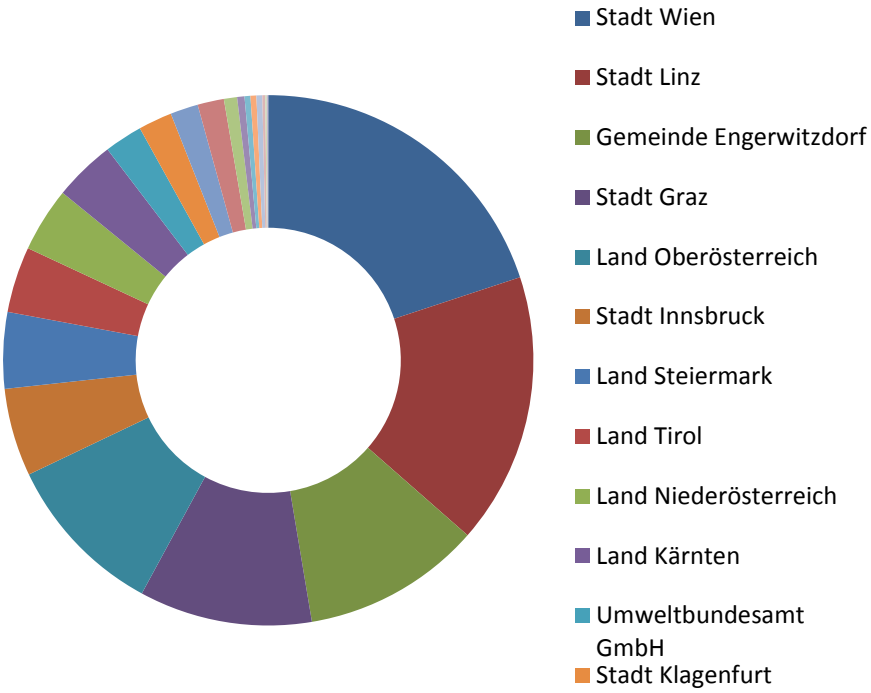


Abbildung 7: Überblick Organisationen

2.5.2 Überblick über verfügbare Datenquellen nach Kategorie

Die zur Verfügung gestellten Datenquellen werden auf der Plattform auch in unterschiedliche Kategorien von Verwaltung und Politik bis hin zu Arbeit eingeteilt. Mehr als die Hälfte der Datenquellen werden in den Bereichen Verwaltung und Politik, Umwelt, Bevölkerung und Geographie und Planung bereitgestellt. Ein Überblick über die unterschiedlichen Kategorien und die Anzahl der Datenquellen ist in Abbildung 8 abgebildet.

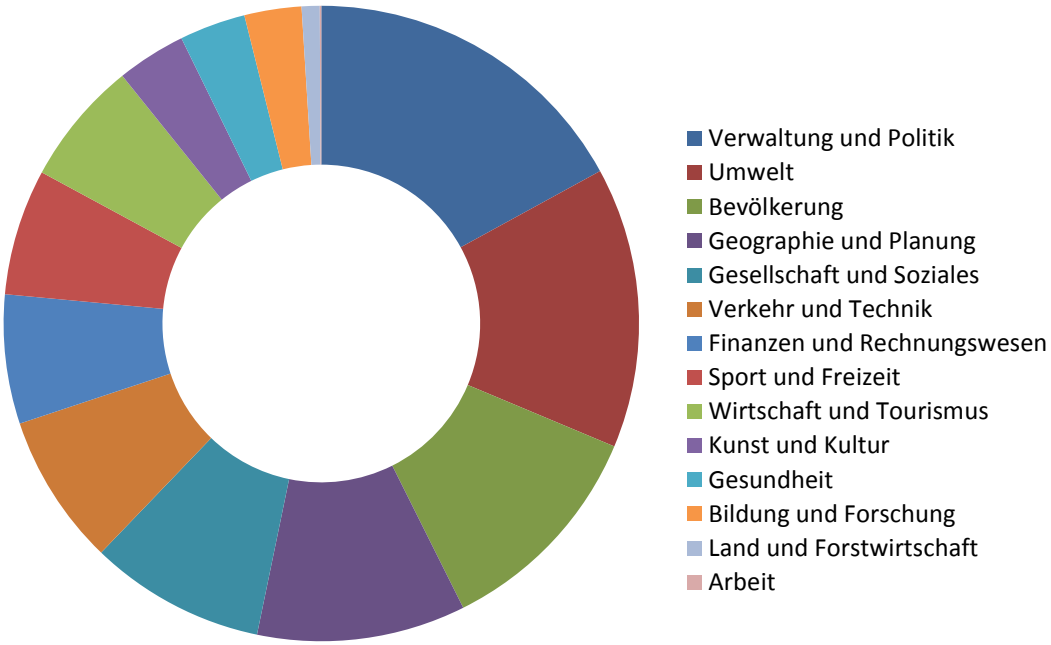


Abbildung 8: Überblick Kategorien

2.5.3 Überblick über verfügbare Datenquellen nach Formaten

Eine Grundvoraussetzung für die sinnvolle Weiterverwendung von öffentlichen Datenquellen ist deren Bereitstellung in offenen Formaten. Dies wird von den meisten Datenquellen berücksichtigt. Eine Aufstellung der am häufigsten verwendeten Formate findet sich in Abbildung 9. Durch diese Zurverfügungstellung der Daten wird die Weiterverwendung in Anwendungen und die Erstellung von neuen innovativen Applikationen erst ermöglicht. Das Portal data.gv.at bietet hier die Möglichkeit, Applikationen zu registrieren, was bisher 174-mal genutzt wurde.

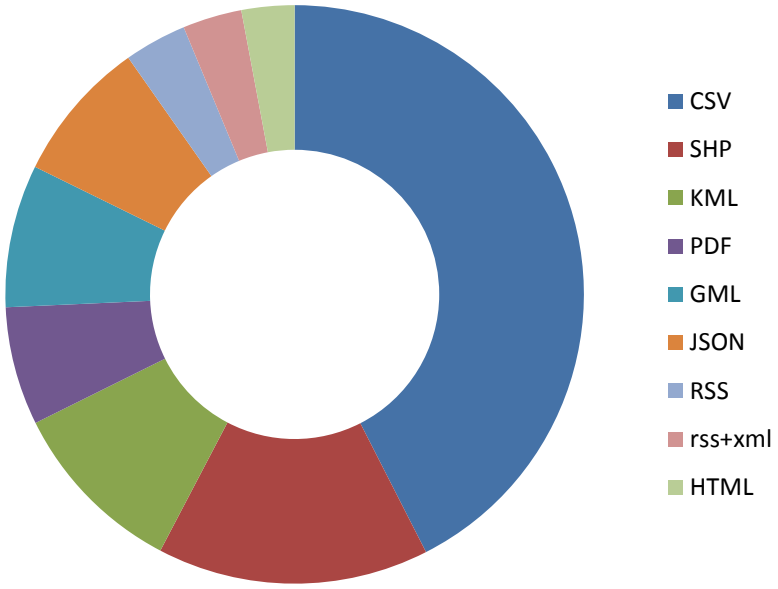


Abbildung 9: Überblick Datenformate

2.5.4 Initiativen, Lizenz und Aussicht

Die Zurverfügungstellung der Daten unter einer Lizenz, die deren Weiterverwendung für Innovationen ermöglicht, ist essenziell. Durch die Verwendung der Creative Commons Namensnennung 3.0 Österreich wird dieses Ziel unter der Voraussetzung der Namensnennung des Autors/der Autorin bzw. des Rechteinhabers/der Rechteinhaberin erreicht.

Die offene und gemeinnützige Plattform open3.at steht für Open Society, Open Government und Open Data und versucht sich als Intermediärin zwischen beteiligten Gruppen zu positionieren und somit diese Bereiche in Österreich voranzutreiben. Die private Initiative opendata.at verfolgt das Ziel, öffentliche Daten in menschen- und maschinenlesbarer Form der Bevölkerung und der Wirtschaft zur Verfügung zu stellen.

Ein wichtiger Teilbereich der öffentlichen Zurverfügungstellung von Daten ist der Bereich offene Verkehrsdaten. Dieser ist derzeit noch stark unterrepräsentiert, birgt aber ein großes Innovationspotenzial, welches derzeit noch nicht vollständig ausgeschöpft wird. Derzeit stehen die Daten der Verkehrsdaten der Wiener Linien und der Linz Linien GmbH unter der Creative Commons Namensnennung 3.0 Österreich Lizenz zur Verfügung. Die Stadt Graz hat kürzlich über eine Pressemitteilung die Zurverfügungstellung von Verkehrsdaten angekündigt. Die gemeinsame österreichische Initiative Verkehrsauskunft Österreich verfolgt das Ziel, das ganze Verkehrsgeschehen in Österreich abzudecken, und wird von mehreren starken PartnernInnen

betrieben. Aktuellen Informationen zu Folge sind derzeit keine Pläne in Richtung Open Data bekannt geworden. Auf Grund einer Anfragebeantwortung von Bundesministerin für Verkehr, Innovation und Technologie (BMVIT, 2014) Doris Bures[42] ist bekannt, dass zwischen den VAO-Betriebspartnern vereinbart wurde, für Forschungszwecke Dritter Testmandanten einzurichten mit deren Hilfe einzelne Teile der Daten von Forschungseinrichtungen verwendet werden können.

Aus strategischer Sicht beinhaltet eine öffentliche Bereitstellung und Nutzbarmachung von Daten ein sehr hohes Potenzial für Innovationen in Österreich, sowohl in der Forschung als auch in der Wirtschaft. Ein wichtiger Schritt hierbei ist aber auch die Verwendung von konsolidierten Formaten, Beachtung von Datenqualität sowie die Aktualisierung von Daten. Ein weit reichendes Konzept in Richtung Open Data kann den österreichischen Forschungs- und Wirtschaftsstandort in den Bereichen IKT, Mobilität und vielen anderen stärken und auch einen internationalen Wettbewerbsvorteil bewirken.
Ein weiteres, wichtiges Portal ist das "Open Data Portal", welches mit 01. Juli 2014 in Betrieb geht. Hierbei handelt es sich um das Schwesternportal von "data.gv.at".

[42] ÖBB sowie VAO Echtzeitdaten und Open Government Data: http://www.parlament.gv.at/PAKT/VHG/BR/AB-BR/AB-BR_02738/index.shtml

3 Markt- und Potenzialanalyse für Big Data

Die in diesem Kapitel dargestellte Markt- und Potenzialanalyse ist dreistufig aufgebaut. In einer ersten Stufe wird ein Überblick über den weltweiten Big Data Markt gegeben. Der Fokus liegt hierbei auf den gegebenen Möglichkeiten auf Grund der aktuellen Marktsituation sowie auf wichtigen Entwicklungen. In einer zweiten Stufe werden der europäische Markt und dessen Potenzial beleuchtet, bevor der eigentliche Fokus in der dritten Stufe auf den österreichischen Big Data Markt gelegt wird. Hierbei werden die jeweiligen Sektoren vorgestellt und deren Herausforderungen, Potenziale und Business Cases analysiert. Herausforderungen wurden auf Basis von Umfragen für den jeweiligen Sektor sowie generellen industriespezifischen Herausforderungen erarbeitet. Potenziale und Use Cases ergeben sich aus den jeweiligen Möglichkeiten, welche weltweit bereits eine erste (erfolgreiche) Umsetzung erfahren haben. Hierbei geht es jedoch oftmals nicht um die in Österreich durchgeführten Projekte (auf diese wird in Kapitel 3 näher eingegangen), sondern vielmehr um generelle Aspekte des Einsatzes von Big Data Lösungen im jeweiligen Wirtschaftszweig.

3.1 ×berblick über den weltweiten Big Data Markt

IDC erwartet, dass sich der weltweite Big Data Markt von 9,8 Milliarden USD im Jahr 2012 auf 32,4 Milliarden USD im Jahr 2017 steigern wird. Das entspricht einer jährlichen Wachstumsrate von 27% (CAGR = Compound Anual Growth Rate). Ende 2013 hatte der Big Data Markt bereits eine Größe von 12,6 Milliarden USD. Anbieter von Big Data Technologien haben hierbei die Möglichkeit, auf allen Ebenen des Technologie-Stacks zu wachsen. Der Technologie-Stack ist im Detail in Kapitel 2.1.3 als U-A-P-M Modell beschrieben.

Der größte relative Anteil am Wachstum liegt in der Kategorie Hardware für Cloud Infrastrukturen mit einer jährlichen Wachstumsrate von 48,7% im Vergleichszeitraum 2012-2017. Im Big Data Stack der Studie ist dies der Ebene "Management" zuzuordnen. Dieser folgt die Kategorie Storage mit einer Wachstumsrate von 37,7%. Der Markt für Big Data Softwarelösungen wächst mit 21,3%, der Services-Bereich mit 27,1%.

Im weltweiten Markt gibt es einige Schlüsseltrends im Big Data Umfeld welche in weiterer Folge auch für den österreichischen Markt sehr relevant sind:

- Wichtige Big Data Kompetenzen sind oftmals nur begrenzt in Anwenderunternehmen vorhanden. Das führt dazu, dass zum einem sehr stark auf Outsourcing-Provider zurückgegriffen werden muss und anderseits, dass oftmals „out-of-the-box“ Lösungen bezogen werden. Bei „out-of-the-box“ Lösungen wird oftmals auf Cloud-Lösungen gesetzt. Um die fehlenden Kenntnisse zu ersetzen, werden viele Unternehmen vermehrt die Automatisierung der IT-Landschaft verstärken. Hierbei werden Cloud-Lösungen eine wesentliche Rolle spielen. Der Wirtschaftsbereich Cloud-Lösungen wird aus diesen Gründen auch deutlich schneller wachsen als On-Premise-Lösungen.
- In den letzten Monaten wurden sehr viele Venture Capitals für Big Data Unternehmen vergeben. Das wird in weiterer Folge zu vermehrten Übernahmen von Big Data Unternehmen durch große Unternehmen führen.
- Das Hadoop Ökosystem wird in naher Zukunft eine sehr bedeutende Rolle spielen. Diese Big Data Lösungen werden jedoch klassische Data Warehousing Lösungen nicht ersetzen.

Vielmehr werden Big Data Lösungen in vielen Unternehmen neben Data Warehousing Systemen als Ergänzung betrieben werden.
- Internet of Things (IoT) und maschinengenerierte Daten werden eine wichtige Rolle im zukünftigen Wachstum von Daten spielen.
- Entscheidungsunterstützungslösungen werden ebenfalls eine wesentlich stärkere Automatisierung erfahren. Diese Lösungen werden in der Folge verstärkt die Rolle des „Wissensmanagers“ ersetzen.

In den jeweiligen Branchen sieht das Bild derzeit folgendermaßen aus:

- Der Einzel- und Großhandel ist unter starkem Leistungsdruck, datengetriebene Anwendungen zu erstellen. Unternehmen, welche diese Anwendungsarten noch nicht verwenden, fallen im globalen oder regionalen Wettbewerb deutlich zurück.
- Im öffentlichen Sektor werden vor allem in Nordamerika derzeit große Investitionen getätigt. Schlüsselministerien waren hierbei einerseits das Bildungswesen U.S. Department of Education mit der Umsetzung von Big Data-getriebener Betrugserkennung im Bereich von Förderungen und andererseits das U.S. Department of the Treasury’s, welches verstärkt Big Data für die Erfassung von Unternehmen und Personen mit schlechter Zahlungsmoral einsetzt.
- Im Energiesektor gibt es erste vorsichtige Schritte in Richtung Big Data. Diese Industrie hinkt jedoch der allgemeinen Entwicklung hinterher.
- Big Data wird bereits stark im Gesundheitsbereich eingesetzt. Hierbei gibt es eine Vielzahl von Anwendungsfällen. Diese reichen von Versicherungen bis hin zu Patientendaten für eine bessere Behandlung und Symptomerkennung.
- In der Öl- und Gasindustrie gibt es bereits sehr ausgereifte Projekte im Big Data Umfeld. Diese Industrie ist sehr weit fortgeschritten und sieht Big Data als wesentliche IT-Richtung für die kommenden Jahre.
- Im landwirtschaftlichen Bereich wird Big Data nicht besonders ausgeprägt eingesetzt. Es gibt hier erste vorsichtige Entwicklungen, diese sind jedoch meist auf sehr wenige Pilotprojekte begrenzt.

3.1.1 Situation in Europa

Der europäische Markt ist noch wesentlich konservativer als der nordamerikanische oder asiatische Markt. Die zentralen Entwicklungen im europäischen Markt sind:

- Das Wachstum im Big Data Umfeld wird im Jahr 2014 bedeutend ansteigen. Erste Projekte konnten von wichtigen Playern im Markt bereits Ende 2013 gewonnen werden. Dieser Trend wird sich im Jahr 2014 fortsetzen.
- Ein Hemmfaktor ist jedoch die Rechtsprechung und Datensicherheit. Dieses häufig diskutierte Thema wird auch 2014 für gehörigen Diskussionsstoff sorgen.
- So genannte „Killer Applications“ werden in den jeweiligen Branchen entstehen.
- Durch Big Data wird eine stärkere Bedeutung für Analytics entstehen. Ebenso wird dieses Thema besser verstanden werden.
- Verschiedene Länder in der europäischen Union werden unterschiedlich auf das Thema Big Data aufspringen. Daher werden die Adaptionsraten auch stark variieren. Besonders stark starten wird das Thema in Nordeuropa.

- Kleine und mittlere Unternehmen werden sich primär auf die Cloud für Big Data Lösungen verlassen. Hier besteht ein hohes Potenzial für Lösungen, welche eine bedeutende Vereinfachung bieten, wie etwa Hadoop als PaaS-Lösung.

3.2 Marktchancen in Österreich

In dieser Kategorie wird der Markt in Österreich dargestellt. Hierbei wird auf die vorhandenen Industrien eingegangen und dargestellt, welche Herausforderungen und Potenziale es gibt. Die jeweiligen Herausforderungen und Potenziale leiten sich primär aus bereits vorhandenen Studien der Projektpartner, öffentlich zugänglichen Studien sowie vorhandenen Fallstudien ab.

3.2.1 Marktüberblick

Der Gesamtmarkt in Österreich war im Jahr 2011 und 2012 noch sehr schwach. Ein verstärktes Wachstum wird in den Jahren 2013-2015 einsetzen. Durch mehrere Gespräche mit verschiedenen Stakeholdern während der Studienerstellung wurde festgestellt, dass oftmals Projekte im Big Data Umfeld Ende 2013 beziehungsweise Anfang 2014 gewonnen wurden. IDC-Recherchen haben ergeben, dass der Markt für Big Data im Jahr 2012 ein Volumen von 18,9 Millionen EUR hatte. Im Jahr 2013 ist dieser moderat mit 21,6% auf 22,98 Millionen EUR angewachsen. Für die kommenden Jahre wird das Wachstum stark beschleunigt, vor allem dadurch, dass die Akzeptanz steigt und das Wissen über die Möglichkeiten der jeweiligen Technologien auch besser wird. Die folgende Tabelle und Abbildung stellen das Wachstum dar.

Jahr	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**CAGR**
Marktvolumen in Millionen EUR	18,90	22,98	31,57	42,78	56,52	72,85	**31,0%**
Wachstum		**21,6%**	**37,4%**	**35,5%**	**32,1%**	**28,9%**	

Tabelle 6: Wachstum von Big Data Technologien in Österreich, IDC

Abbildung 10: Wachstum von Big Data Technologien in Österreich, IDC

Hierbei ist klar zu erkennen, dass die Technologien im Big Data Stack bis ins Jahr 2013 nur moderat gewachsen sind. Ab dem Jahr 2013 steigt das Wachstum nun stark an. In anderen Ländern, wie beispielsweise den USA, ist das Wachstum bereits wesentlich früher gestartet. Diese abwartende Haltung wurde auch in anderen Technologien beobachtet. Ein Beispiel hierfür ist etwa Cloud Computing. Eine Folge davon ist, dass das Wachstum in den folgenden Jahren daher wesentlich höher ausfällt als im weltweiten Vergleich.

Das Wachstum bedeutet jedoch nicht unbedingt, dass österreichische Unternehmen dadurch stark davon profitieren. Dadurch, dass es kaum österreichische Unternehmen in diesem Sektor gibt, die eine bedeutende Rolle spielen, wandert ein großer Teil der Wertschöpfung ins Ausland ab. Bedeutende Unternehmen im Big Data Umfeld betreiben derzeit oftmals keine dedizierte Niederlassung in Österreich. Die Kundenbetreuung und der Verkauf wird vielfach aus Deutschland abgehandelt, was in weiterer Folge bedeutet, dass die Wertschöpfung nicht in Österreich stattfindet. Nach und nach werden auch österreichische Unternehmen auf Technologien im Big Data Stack wechseln, wobei hierbei der Markt bereits eine gewisse Dynamik entwickelt hat. Internationale Unternehmen mit einer Niederlassung in Österreich werden bereits früher ein Wachstum vorweisen können und dieser Bereich wird stärker wachsen. Die Entwicklung der Zahlen wurde durch Einzelgespräche mit Vertretern aus diesen Unternehmenskategorien modelliert.

Für diese Prognose wurde primär die Ist-Situation evaluiert. Durch geziehlte Maßnahmen kann hier jedoch gegengesteuert werden. Diese werden im Kapitel 5 dargestellt.

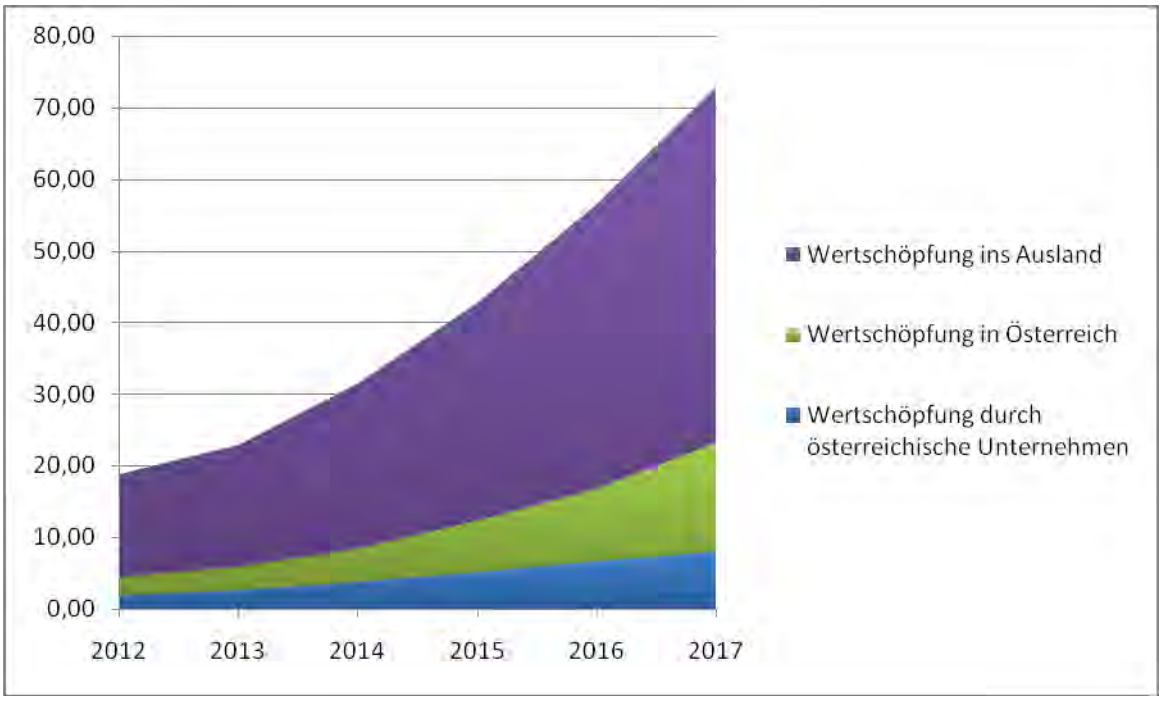


Abbildung 11: Wertschöpfung in Österreich

3.2.2 Akzeptanz innerhalb Österreichs

Eine jährlich durchgeführte Studie der IDC Österreich gibt weitere Aufschlüsse über dieses Thema. Im November 2012 wurde erstmals die Meinung der IT-Entscheidungsträger in Österreich evaluiert. Diese Studie wurde im November 2013 wieder durchgeführt. Das Sample hierfür war 150. Beeindruckend ist der Anteil jener, die angaben, dass Big Data nicht diskutiert wird. War dieser Anteil 2012 noch bei 60%, so ging dieser im Jahr 2013 auf unter 40% zurück. Dies ging großteils zugunsten derjenigen, die Big Data im Unternehmenseinsatz derzeit diskutieren. Die Anzahl jener RespondentInnen, die bereits Projekte umsetzen oder planen, ist ebenfalls gestiegen. Derzeit besteht aufgrund dieser Tatsachen ein gesteigertes Beratungspotenzial. Dies wird in Abbildung 12: Akzeptanz von Big Data Lösungen in Österreich dargestellt.

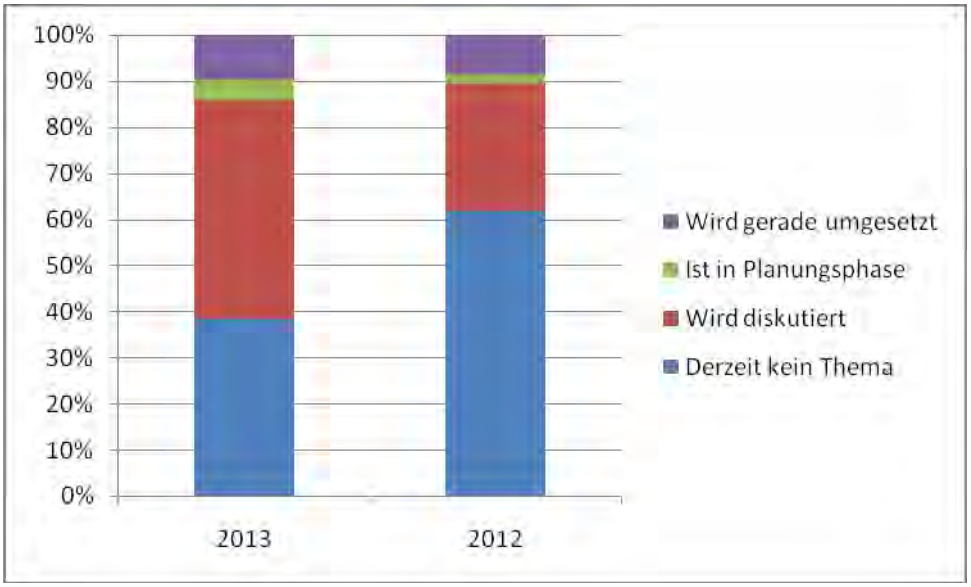


Abbildung 12: Akzeptanz von Big Data Lösungen in Österreich

Besonders interessant gestaltet sich auch die Unterscheidung nach Unternehmensgröße. Big Data ist vor allem bei großen Unternehmen interessant, wohingegen ein Großteil der KMU's das Thema noch sehr skeptisch sieht. Hierbei wird oftmals argumentiert, dass Big Data für KMU's aufgrund der „Größe" nicht bewältigbar ist. Dies wird in Abbildung 13: Big Data nach Unternehmensgröße dargestellt.

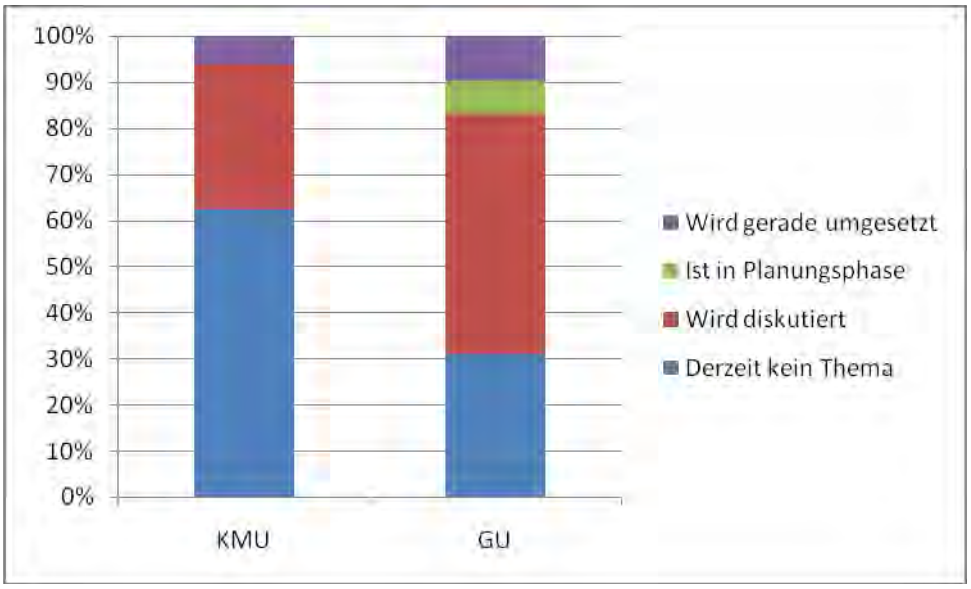


Abbildung 13: Big Data nach Unternehmensgröße

Hemmfaktoren bestehen vor allem aufgrund der Novität der Technologie. Viele Unternehmen haben Angst vor der Komplexität der Anwendungen und fürchten, dass das interne Know-how nicht ausreicht. Auch oder wegen des Einsatzes von Cloud-Anwendungen für Big Data sind Hemmfaktoren für Skalierung und Bereitstellung stark zurückgegangen, wobei die Angst hinsichtlich der Sicherheit der Daten gestiegen ist. Wenn AnwenderInnen mit eigenen Lösungen erfolgreich sein möchten, so wird es notwendig sein, eine wesentliche Vertrauensbasis zu potenziellen KundInnen zu schaffen. Dies wird in Abbildung 14: Hemmfaktoren beim Einsatz von Big Data Lösungen dargestellt.

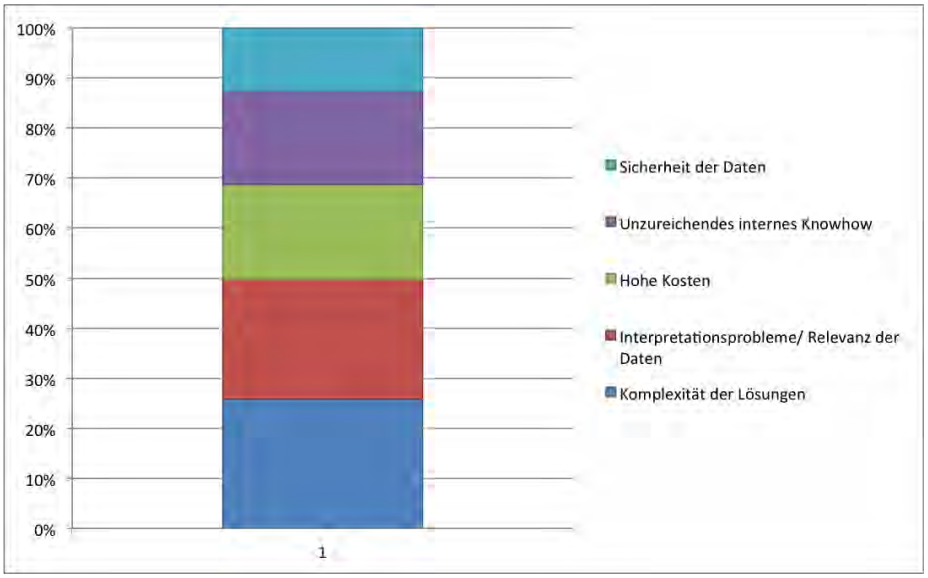


Abbildung 14: Hemmfaktoren beim Einsatz von Big Data Lösungen

Heimische Unternehmen erwarten sich allgemein bessere Datenanalysen, um bestehende Analysen zu verbessern. Ebenso wichtig ist die Erkennung von Kaufmustern. Dies kann vor allem im Retail-Bereich zu Vorteilen führen. Dies wird in Abbildung 15: Erwartete Einflüsse von Big Data dargestellt.

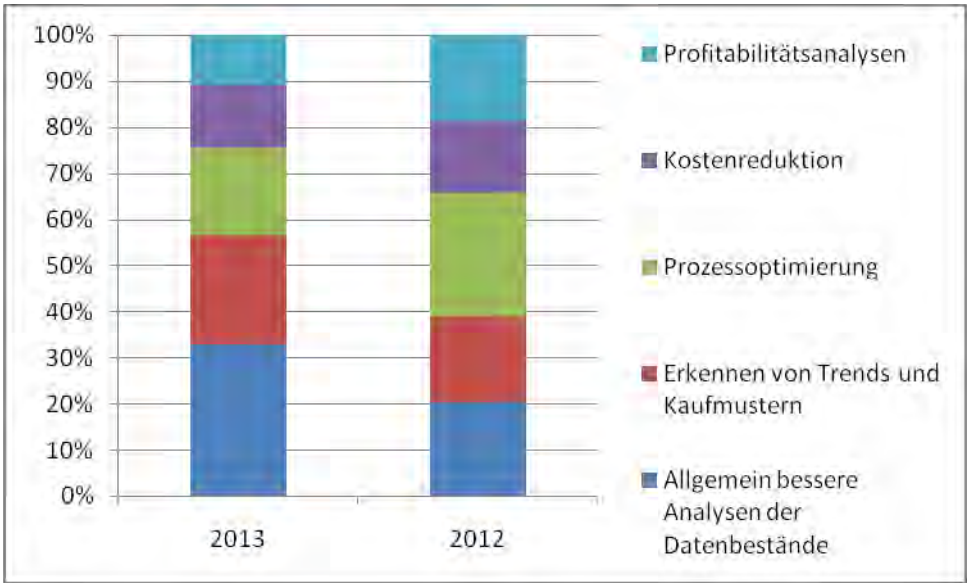


Abbildung 15: Erwartete Einflüsse von Big Data

Für 46% der befragten Unternehmen sind der Aufbau des Know-hows und das Verstehen der Prozesse jener Bereich, in dem der größte Handlungsbedarf besteht. Dahinter folgen der Aufbau der IT-Umgebung für entsprechende Datenspeicherung und Analyse - 21% - und die Auswahl der richtigen Software - 13%. Dies wird in Abbildung 16: Was ist in heimischen Unternehmen notwendig, um Big Data einzuführen? Dargestellt.

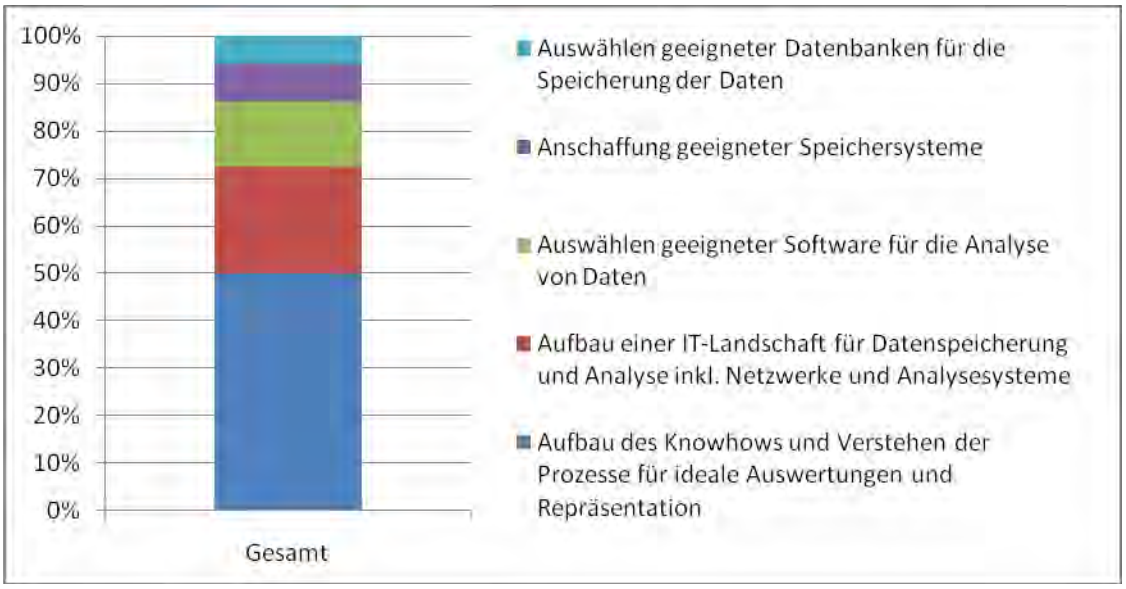


Abbildung 16: Was ist in heimischen Unternehmen notwendig, um Big Data einzuführen?

Aktuell befindet sich Big Data in einem initialen Stadium. Die Einsatzentscheidung ist oftmals schon gefallen, es mangelt Endusern jedoch noch am Know-how bzgl. Anbietern, Verfahren und Werkzeugen. Diesen Umstand gilt es seitens der AnbieterInnen zu nutzen und Unternehmen über

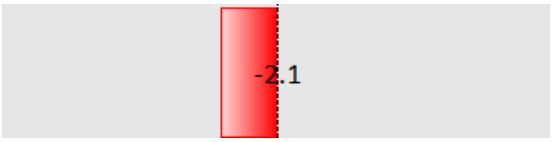
zur Verfügung stehende Tools und Werkzeuge im Zusammenhang mit Big Data zu informieren – AnbieterInnen können sich so frühzeitig Wettbewerbsvorteile verschaffen.

Big Data ist eine abteilungsübergreifende Technologie. Das spiegelt sich auch in den Ergebnissen dieser Befragung wider und ist ein wichtiger Einflussfaktor für Enduser-Entscheidungen. Sehr bedeutend für Big Data ist die Analyse von unstrukturierten Daten, die einen wesentlichen Einfluss auf die IT-Landschaft eines Unternehmens haben.

3.2.3 Standortmeinungen internationaler Unternehmen

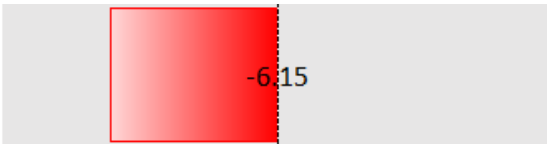
Im Zeitraum der Studie wurden viele Einzelgespräche mit bedeutenden Big Data Unternehmen geführt, welche keine Niederlassung in Österreich haben, aber einen aktiven Verkauf im Land betreiben. Hierbei wurden die Fragen vor allem auf den Standort und die Wettbewerbsfähigkeit gerichtet. Die Relevanz ist hierbei dafür interessant, ob und wie schnell Investitionen von außerhalb getätigt werden. Ferner soll dies eine „Außenansicht" auf den Markt geben, also Antwort auf die Frage liefern, wie Österreich von außen hinsichtlich Big Data gesehen wird und wie der heimische Markt international aufgestellt ist. Hierfür wurden rund 20 Unternehmen im Big Data Umfeld befragt. Die Befragten stammten primär aus Deutschland und England. Die Skala reicht von -10 (extrem negativ) bis +10 (extrem positiv).

Wie schätzen Sie das Potential des österreichischen Marktes hinsichtlich Big Data ein?



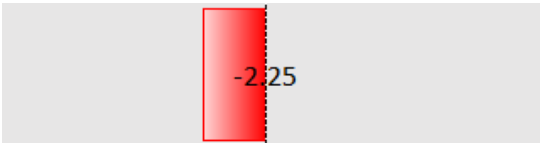
Die erste Frage wurde hinsichtlich des allgemeinen Potenzials evaluiert. Hierbei wurde nachgefragt, wie der österreichische Markt aktuell eingeschätzt wird und wie sich dieser potenziell zukünftig entwickeln kann. Hierbei haben die RespondentInnen angegeben, dass der heimische Markt ein leicht negatives Potenzial gegenüber anderen europäischen Ländern hat. Besonders von Relevanz wurde bei den Befragten der englische, französische, deutsche, polnische und nordische Markt angegeben. Österreich steht bei den meisten Unternehmen nicht auf den ersten Rängen, wenn es um zukünftige Investitionen und Ausbau der Tätigkeiten geht.

Schützen Sie, dass Sie in den nächsten 2 Jahren einen Standort in Österreich haben werden?



In der zweiten Frage wurde evaluiert, in welchem Zeitraum ein Unternehmen eine aktive Niederlassung in Österreich haben wird. Hierbei gab der Großteil an, dass für die nächsten zwei Jahre keine Niederlassung in Österreich angedacht sei und der Verkauf aus anderen Ländern gesteuert wird. Damit besteht für Österreich die Gefahr, dass Know-how abwandert und die Wertschöpfung im Bereich ebenfalls Österreich verlassen wird. Als Grund wurde oftmals die konservative Einstellung in Österreich genannt, wodurch Unternehmen wenig bis kaum in moderne Technologien investieren und dadurch auch die Entwicklungschancen heimischer DienstleisterInnen und Unternehmen gering gehalten sind. In Ländern, in denen die Technologie bereits besser angenommen wird, entwickelt sich ein Ökosystem rund um Big Data, wodurch Know-how aufgebaut und ein Wettbewerbsvorteil erarbeitet wird.

Im europäischen Vergleich: wie schätzen Sie Österreich ein?

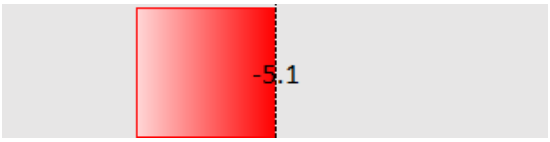


Ähnlich gestaltet sich die Frage nach dem europäischen Vergleich. Hier schneidet Österreich negativ ab. Viele der befragten Personen gaben an, dass in anderen Ländern bereits wesentlich mehr in diesem Thema erledigt wird. Außerdem seien IT-EntscheiderInnen in anderen Ländern bei Weitem nicht so verschlossen, wenn es um dieses Thema ginge. In Österreich sind CIO's oftmals nicht über die Möglichkeiten informiert, sondern berufen sich lediglich auf Datenschutz.

Wie bewerten Sie die Innovationsfreudigkeit österreichischer Unternehmen, neue Technologien (vor allem Big Data) anzunehmen?

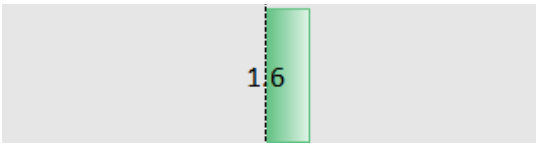
Diese Frage hatte bereits einen großen Einfluss auf andere Fragen. Vielfach wurde angegeben, dass heimische Unternehmen kaum innovationsfreudig sind. Der Vergleich wurde hierbei mit anderen europäischen Ländern angestellt. Hierbei schnitt Österreich im Prinzip fatal ab. Bemängelt wurde oftmals, dass Verantwortliche im Unternehmen sich gar nicht erst mit dem Thema „Big Data" auseinandersetzen wollen und es einem klassischen BI Thema gleichstellen.

Wie schätzen Sie die Chancen ein, dass österreichische Unternehmen eine führende Rolle in internationalen Märkten spielen?



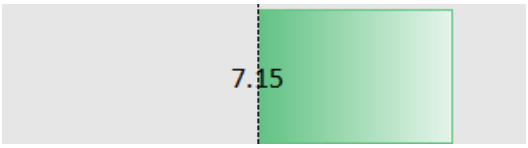
Hierbei wurde der Fokus vor allem auf die Frage gerichtet, ob und wie sich österreichische Unternehmen am internationalen Markt etablieren können. Der Grundtenor lautete, dass bereits vieles an Innovation in anderen Ländern erfolgt ist und es dadurch sehr schwierig ist, diesen Rückstand noch aufzuholen. Prinzipiell sei es jedoch möglich, wobei es hierfür erheblichen Aufwandes bedarf. Ob dieser aufgebracht werden kann, wurde vielfach mit „fraglich" abgetan.

Wie schätzen Sie das Ausbildungsniveau in Österreich für zukünftige Investitionen Ihres Unternehmens in den Standort ein?



Leich positiv wird hingegen die Ausbildungssituation bewertet. So gibt es laut den Befragten in den jeweiligen Unternehmen einige ÖsterreicherInnen mit guten Qualifikationen. Diese haben jedoch vielfach das Land verlassen. Die Ausbildungssituation wird generell als etwas besser als mittelmäßig bezeichnet. Es ist hier noch deutlich Luft nach oben, jedoch muss man sich als ÖsterreicherIn auch nicht gänzlich verstecken. Generell wurde nahegelegt, dass die Bildungseinrichtungen aktiver auf das Thema zugehen sollen.

Wie schätzen Sie die Auswirkungen gezielter Investitionen auf den Sektor?



Werden in Österreich gezielt Investitionen in den Bereich „Big Data“ unternommen, so wird allgemein angenommen, dass sich die Situation wesentlich verbessert.

3.2.4 Branchen in Österreich

In diesem Kapitel werden die jeweiligen Branchen in Österreich beschrieben. Hierbei wird die Branche beschrieben und in weiterer Folge die jeweiligen Herausforderungen und Potenziale für Anwendungen im Big Data Umfeld. Hierbei wird primär von internationalen Case Studies ausgegangen und diese für den Standort Österreich diskutiert. Für die Klassifizierung wurde der ÖNACE 2008-Standard verwendet[43].

3.2.4.1 Land- und Forstwirtschaft, Fischerei

Die Land- und Forstwirtschaft hat seit Jahrzehnten einen rückläufigen Anteil einerseits beim Beitrag zur nationalen Bruttowertschöpfung und andererseits bei den in der Land- und Forstwirtschaft beschäftigten Personen. Auch die Anzahl der land- und forstwirtschaftlichen Betriebe geht stetig zurück. Nichtsdestoweniger hat die Land- und Forstwirtschaft gerade im ländlichen Raum einen erheblichen Einfluss auf andere, vor- und nachgelagerte Wirtschaftsbereiche wie z.B. Landmaschinenindustrie, Nahrungsmittelindustrie, Bau, Handel, Dienstleistungsbereich und zeichnet sich maßgeblich für die Landschaftspflege verantwortlich, wovon wiederum die Tourismuswirtschaft profitiert. Zu den bedeutendsten Unternehmen in dieser Branche gehören etwa der Waldverband Steiermark, die Styriabrid GesmbH sowie die Österreichische Bundesforste AG.

Herausforderungen	Die Land- und Forstwirtschaft und Fischerei sind sehr traditionelle Branchen, welche oft im kleinen Umfeld (Einzelpersonen oder Familien) und subventioniert betrieben werden. Innovationen finden hier oftmals nicht so schnell Zuspruch, was natürlich auch an der Grundlage dieser Branchen liegt. Eine zentrale Herausforderung liegt darin, die Technologien für diese Branchen überhaupt erst interessant zu machen. Wesentliche Herausforderungen bestehen in der wachsenden Weltbevölkerung, welche auch vor Österreich keinen Halt macht. Hierbei stellt sich die Frage, wie diese mit der Landwirtschaft und Fischerei effektiv ernährt werden kann.
Potenziale und Use Cases	In der Land- und Forstwirtschaft wie auch der Fischerei sind die Potenziale von Big Data im Vergleich zu anderen Branchen begrenzt. In relevanten Umfragen hat sich herausgestellt, dass kaum relevante Einsatzgebiete vorkommen. Gewisse Potenziale bestehen jedoch in der Landschaftserfassung und

[43] http://www.statistik.at/KDBWeb/kdb_VersionAuswahl.do

der Auswertung von Sensordaten. Das Unternehmen John Deere setzt seit 2012 verstärkt auf Produkte, welche es den landwirtschaftlich Bediensteten ermöglichen, die Erträge zu optimieren. Hierbei werden Einsätze auf Basis verschiedener Datenquellen wie dem Wetter, der Beschaffenheit der Landschaft sowie der zu bearbeitenden Elemente (z.B. Getreideart) optimiert. Dadurch kann der Energieverbrauch von Landwirtschaftlichen Geräten optimiert und wertvolle Zeit eingespart werden. Wesentliche Optimierungen können auch für die ertragreichste Anpflanzung von z.B. Getreidesorten durchgeführt werden.

3.2.4.2 Bergbau und Gewinnung von Steinen und Erden

Im österreichischen Bergbau werden derzeit von mehr als 5.000 ArbeitnehmerInnen in mehr als 1.300 Betriebsstätten jährlich ca. 120 Mio. Tonnen feste mineralische Rohstoffe gewonnen sowie knapp 1 Mio. Tonnen Rohöl und mehr als 1,5 Mrd. m³ Erdgas gefördert. Dennoch ist diese Branche in Österreich traditionell von eher geringer Bedeutung. Zu den wichtigsten Unternehmen in dieser Branche zählen die Salinen Austria AG, die Montanwerke Brixlegg AG, sowie die voestalpine Stahl GmbH.

Herausforderungen

Eine wesentliche Herausforderung in dieser Branche ist die geringe Bedeutung in Österreich. Dadurch ist es bestimmt nicht einfach, Investitionen zu tätigen. In dieser Branche gibt es jedoch eine große Anzahl an wartungsintensiven Geräten, welche vor allem durch Datenanalysen verbessert werden können.

Potenziale und Use Cases

In dieser Branche bestehen gewisse Möglichkeiten für Optimierungstechniken in den jeweiligen Abbauverfahren. Durch diese wird es möglich, effizienter zu produzieren.

Ein Use Case ist hierbei das Forschungsprojekt von Joanneum Research, wo mithilfe von adaptiven Echtzeitanalysen von Bohrsystemen die Stillstandszeiten minimiert werden konnten.

Vor allem die Echtzeitüberwachung der jeweiligen Systeme bringt die Möglichkeit, Materialverschleiß zu reduzieren und damit die Kosten zu senken. Ferner können die Intervallzeiten wesentlich verbessert werden, indem mehrere Geräte, Maschinen oder Fahrzeuge durch IT-Systeme überwacht und laufend auf Probleme analysiert werden.

3.2.4.3 Herstellung von Waren

Etwa 30.000 Unternehmen in Österreich sind im „Sachgüterbereich“ genannten Wirtschaftsbereich tätig. Der Schwerpunkt der Aktivitäten liegt naturgemäß im Bereich Herstellung von Waren. Hier waren zuletzt mehr als 25.000 Unternehmen tätig, die mit über 600.000 Beschäftigten Umsatzerlöse von mehr als 170 Mrd. Euro erzielten. Bedeutende Repräsentanten dieser Branche sind etwa Jungbunzlauer Austria, Sony DADC Austria, oder AGRANA Zucker.

Herausforderungen	Der Wettbewerbsdruck in dieser Branche ist vor allem aufgrund des starken Wachstums in Asien geprägt. Billigere Produktionskosten bringen den Standort Österreich in Gefahr. Dieser Sektor ist vor allem investitionsintensiv. Neue Produktionsanlagen benötigen Kapital, aber auch die jeweiligen Sachgüter, welche produziert werden, haben einen hohen Investitionsbedarf. Gefahr entsteht zudem durch Fehlproduktionen
Potenziale und Use Cases	Wesentliche Vorteile ergeben sich durch die Produktionsverbesserung durch Echtzeit-Analysen. Wenn dadurch die Produktionsfehler gesenkt werden können, senken sich auch die Einzelkosten, was in weiterer Folge wesentliche Wettbewerbsvorteile bringt. Weitere wichtige Vorteile ergeben sich aus der Forschung: durch eine ständige Optimierung der Produkte, welche vor allem durch Datenanalysen vorangetrieben werden, können die Produktionskosten weiter gesenkt werden. Wesentliche Potenziale ergeben sich auch in der Produktion an sich: durch Echtzeitdaten der Auftragsvolumina können Über- bzw. Unterproduktionen verhindert werden.

3.2.4.4 Energieversorgung

Der Wirtschaftszweig Energieversorgung umfasst die Elektrizitäts-, Gas-, Wärme- und Warmwasserversorgung durch ein fest installiertes Netz von Strom- bzw. Rohrleitungen. Inkludiert ist sowohl die Versorgung von Industrie- und Gewerbegebieten als auch von Wohn- und anderen Gebäuden. Unter diesen Wirtschaftszweig fällt daher auch der Betrieb von Anlagen, die Elektrizität oder Gas erzeugen und verteilen bzw. deren Erzeugung und Verteilung überwachen. Ebenfalls eingeschlossen ist die Wärme- und Kälteversorgung. In Österreich umfasst diese Branche knapp 2.000 Unternehmen mit rund 29.000 Beschäftigten. An bedeutenden Unternehmen sind hier unter anderen zu nennen: Verbund AG, STEWEAG-STEG, EVN Energievertrieb GmbH oder KELAG.

Herausforderungen	Der ständig steigende Energieverbrauch ist eine der zentralen Herausforderungen in dieser Branche. Zukünftige Herausforderungen umfassen die steigende Anzahl an Fahrzeugen, welche mit Elektrizität betrieben werden, aber auch die Orientierung an erneuerbaren Energien. Eine zentrale Herausforderung für die IT sind die SmartMeter, welche bereits seit geraumer Zeit diskutiert werden.
Potenziale und Use Cases	Erhebliche Potenziale und Use Cases ergeben sich in dem aufstrebenden Bereich Smart Cities mit dem Ziel des intelligenten und nachhaltigen Umgangs mit Ressourcen, welcher sich erheblich auf die Lebensqualität und die Energieeffizienz auswirken kann. Durch die immer stärkere Vernetzung und Automatisierung der Energiewirtschaft können Energienetze intelligent gesteuert werden und der Einsatz von

Big Data Technologien kann hier als Innovationstreiber für die Optimierung der Energienetze und die Verteilung und effiziente Nutzung und Preisgestaltung in Zusammenhang mit Smart Meters gesehen werden. Der verstärkte Einsatz von alternativen Kleinkraftwerken im Solar- und Windbereich und deren starke Bestückung mit Echtzeitsensoren führt zu Smart Grids, welche ebenfalls von Big Data Technologien profitieren und deren Weiterentwicklung treiben können.
In der Auffindung von Rohstoffen (wie etwa Erdgas oder Erdöl) kann durch den verstärkten Einsatz von Echtzeitsensoren und darauf basierenden Datenanalysen ebenfalls eine Produktivitätsverbesserung, eine Effizienzsteigerung sowie eine Vorabanalyse von zu untersuchenden Gebieten erreicht werden.

3.2.4.5 Wasserversorgung, Abwasser- und Abfallentsorgung und Beseitigung von Umweltverschmutzungen

In dieser Branche sind rund 2.000 Unternehmen mit insgesamt etwa 19.000 Beschäftigten erfasst. Insgesamt wurden zuletzt von diesen Unternehmen Umsatzerlöse in Höhe von 4,8 Mrd. Euro erwirtschaftet. Typische Vertreter dieser Branche sind etwa EVN Wasser GesmbH, oder auch Aqua Engineering GmbH.

Herausforderungen	Umweltschutz ist eine wesentliche Herausforderung der westlichen Welt geworden. Effizienz ist in dieser Branche eine wesentliche Herausforderung, welche mit Big Data adressiert werden kann.
Potenziale und Use Cases	Mithilfe von Datenanalysen kann gemessen werden, wie sich die Wasserqualität in bestimmten Regionen innerhalb Österreichs verändert hat. Durch diese Analysen können Warnungen und/oder Handlungsalternativen ausgegeben werden. Potenziale bestehen aber auch in anderen Bereichen, wie etwa der Entwicklung von Wäldern und Siedlungen und deren Einfluss auf die Umwelt. Ein weiterer wichtiger Punkt ist die Effizienzsteigerung; hierbei wird es möglich, durch gezielte Datenanalysen eine schnellere Durchlaufzeit bei der Abfallentsorgung zu erreichen.

3.2.4.6 Bau

Das Bauwesen ist eine sehr traditionelle Säule der heimischen Wirtschaft und bietet rund 279.000 Personen in Österreich einen Arbeitsplatz. Die rund 31.600 Unternehmen sind vorwiegend kleinbetrieblich strukturiert: Mehr als drei Viertel der Unternehmen beschäftigten traditionell weniger als 10 Personen. Dominante Unternehmen in dieser Branche sind etwa die Porr Bau GmbH oder die Strabag AG.

Herausforderungen

Eine wesentliche Herausforderung besteht im Aufbau der Industrie. Da diese Industrie sehr stark auf Kleinunternehmen setzt, ist die IT hier kaum beziehungsweise gar nicht entwickelt. Es besteht auch wenig bis gar kein Potenzial, dies zu erreichen.

Potenziale und Use Cases

Potenziale bestehen in dieser Industrie primär bei großen Unternehmen des Sektors. In Großprojekten fallen eine Menge an Daten an und um diese Daten effektiv abzulegen und effizient zu verwerten, bedarf es gut funktionierender Systeme. Hierdurch ist es möglich, die Sicherheit von Bauwerken und Gebäuden durch Analysen zu erhöhen beziehungsweise Überwachungssysteme einzubauen, welche die Sicherheit langfristig gewährleisten.

3.2.4.7 Handel, Instandhaltung und Reparatur von Kraftfahrzeugen

Der Handel ist ein wichtiger Wirtschaftszweig im österreichischen Dienstleistungssektor, der ca. 3,7 Mio. österreichische Haushalte versorgt. In Österreich erwirtschafteten ca. 75.000 Handelsunternehmen Umsatzerlöse im Wert von über 240 Mrd. Euro und beschäftigten rund 630.000 Personen, wobei mehr als die Hälfte der ArbeitnehmerInnen im Einzelhandel zu finden sind, der mit rund 40.000 Unternehmen und 350.000 Beschäftigten der größte Arbeitgeber in diesem Bereich war. Zu den bekanntesten Branchenrepräsentanten gehören die SPAR Warenhandels-AG , Herba Chemosan und REWE.

Herausforderungen

KundInnenbindung ist ein zentrales Thema im Handel. Die einzelnen Marktteilnehmer kämpfen besonders intensiv um die jeweiligen KundInnen und wollen diese stärker an das jeweilige Unternehmen binden. Ferner ist es notwendig, das Produktsegment ständig anzupassen.

Potenziale und Use Cases

Im Handel gibt es eine ganze Reihe an Use Cases und Potenziale. Durch KundInnenbindungsprogramme wird der einzelne Kunde oftmals sehr genau „durchleuchtbar“. Einige Handelsunternehmen haben jedoch auch bewiesen, dass Sie ohne personenbezogene Daten gute KundInnenbindungsprogramme ermöglichen können. Besonders von Interesse im Handel ist auch eine Echtzeitüberwachung – wie etwa der Kassensysteme und die Überwachung von jeweiligen Aktionen und deren Leistung in individuellen Filialen. Durch besseres Wissen über den Standort durch Integration von demographischen Daten und Analyse des bestehenden KundInnenstocks kann das Sortiment laufend angepasst und verbessert werden.

Durch Verkaufsanalysen können Verkaufsmuster erstellt werden. So hat ein italienisches Handelsunternehmen herausgefunden, dass gemeinsam mit Slips auch oft Kinderspielzeug gekauft wird. Dadurch kann dies nun effizienter Angeboten werden. In den USA wurde ferner herausgefunden, dass die Bierkäufe mit Windelkäufen

zusammenhängen. Dadurch wurden diese beiden Produktkategorien näher platziert.

3.2.4.8 Verkehr und Lagerei

Transport und Verkehr zählt derzeit über 35.000 Mitglieder in Österreich, beschäftigt rund 200.000 MitarbeiterInnen und erwirtschaftet Erlöse von über 40 Milliarden Euro (Wirtschaftskammer Österreich, 2013). Die Sparte kann in Güterverkehr und Personenverkehr aufgeteilt werden, wobei in beiden hohe Potenziale für Big Data Technologien liegen. Das gesamte Transportaufkommen beträgt gegenwärtig in Österreich rund 450 Mio. Tonnen. Die meisten Güter werden in Österreich mit Straßengüterfahrzeugen befördert, gefolgt von Schiene, Schiff und dem Luftverkehr.

Herausforderungen

Urbanisierung, erhöhtes Transportaufkommen, Klimawandel und Rohstoffknappheit erfordern hier neue Ansätze, um sicherer, umweltfreundlicher und effizienter handeln zu können. Besondere Herausforderungen sind hier vor allem im Zusammenhang mit der effizienten Nutzung vorhandener Infrastrukturen zu erwähnen. Ein besonderes Ziel hier ist es, Maßnahmen zu setzen, die eine bessere Nutzung der vorhandenen Infrastruktur fördern. Wobei das hier umweltfreundlicher, schneller, energieeffizienter oder auch ein Mix aus allen diesen Aspekten sein kann. Diese Maßnahmen können auf der Infrastruktur selbst (bspw. neuer Straßenbelag für mehr Sicherheit), an den Transportmitteln (bspw. Elektromobilität und Leichtmetallkomponenten für Reduzierung der Energiekosten) oder an der Nutzung des multi-modalen Verkehrssystems (bspw. bessere multi-modale Umverteilung des Verkehrs zwecks Effizienz) durchgeführt werden. In diesem Zusammenhang spielen Informationen über die tatsächliche Nutzung der Infrastruktur aus unterschiedlichen Datenquellen zwecks Überwachungs- und Optimierungsaufgaben eine besondere Rolle.

Potenziale und Use Cases

Durch bessere Datenanalysen kann die Nutzung der Infrastruktur analysiert, bewertet und optimiert werden. In der Transportlogistik können hier bspw. Krankentransporte abhängig vom aktuellen Verkehrszustand und Nachfrage so platziert werden, dass diese jederzeit innerhalb eines definierten Zeitraums das ganze Einzugsgebiet versorgen können. Ein weiteres Beispiel ist die mobilfunkbasierte Nutzung über Smartphones und Apps. Diese Informationen können bspw. herangezogen werden, um detaillierte Informationen über die individuelle Mobilitätsnachfrage und Nutzungscharakteristiken zu erhalten, um einerseits personalisierte Informationsdienste zur Verfügung zu stellen, aber auch Informationen über die gesamte Verkehrsnachfrage zu erhalten. Diese wiederum kann genutzt werden, um gezielte verkehrspolitische Maßnahmen zu setzen (z.B. Attraktivierung des öffentlichen Verkehrs durch Anpassung von Fahrplänen)

3.2.4.9 Beherbergung und Gastronomie

Der Bereich Beherbergung und Gastronomie umfasst rund 44.000 Unternehmen mit fast 270.000 Beschäftigten, die ca. 490 Mio. Arbeitsstunden leisteten. Allein in der Gastronomie waren rund 164.000 Personen im Jahresdurchschnitt tätig, während in der Beherbergung im Jahresdurchschnitt nur rund 106.000 Personen beschäftigt waren. Neben den klassischen großen Hotelketten (z.B. Sheraton, Ritz) gehören hier auch Österreichs Thermenressorts (z.B. Tatzmannsdorf) zu den größten Anbietern.

Herausforderungen

Die Gastronomie ist ein wesentliches Standbein der österreichischen Wirtschaft, wenngleich der Wettbewerb hier sehr stark ist. Um sich vom Mitbewerb abzusetzen, müssen ständig neue Marktlücken entdeckt werden.

Potenziale und Use Cases

Ebenso wie im Handel ist die KundInnenbindung hierbei sehr wichtig. Durch bessere Informationen über die Präferenzen der KundInnen kann damit besser auf deren Bedürfnisse eingegangen werden. Somit kann auch das Marketing zielgerichteter erfolgen. Durch die Analyse von Veränderungen der BestandskundInnen kann man das eigene Profil besser schärfen und auf Veränderungen in der KundInnenbasis reagieren. Hier entwickeln sich international in den letzten Jahren stark auf Big Data Analysen basierende neue Geschäftsmodelle und Unternehmen für die individuelle Auswahl und für bedürfnisgerechtes Marketing von Beherbergungsangeboten. In dieser Branche besteht nicht nur großes Potenzial, sondern auch eine große Notwendigkeit innovative Big Data Ansätze für KundInnenbindung, die Gewinnung neuer KundInnen und die Erschließung von neuen Märkten einzusetzen.

3.2.4.10 Information- und Kommunikation

Die Unternehmen der Informations- und Kommunikationsbranche sind ein wesentlicher Zukunftsfaktor für den wirtschaftlichen Erfolg und die Sicherung des Wirtschaftsstandortes Österreich. Diese Branche umfasst u.a. das Verlagswesen, die Herstellung von Filmen und von Tonaufnahmen sowie das Verlegen von Musik, die Herstellung und Ausstrahlung von Fernseh- und Hörfunkprogrammen, die Telekommunikation, Dienstleistungen der Informationstechnologie sowie sonstige Informationsdienstleistungen. Zweifellos ist der ORF der bedeutendste Vertreter in dieser Branche.

Herausforderungen

Diese Branche ist sehr vielfältig zusammengesetzt. Besondere Herausforderungen bestehen für die Informationstechnologie, da Big Data noch nicht so stark in Österreich angenommen wird. Diese Unternehmen können sich auch folglich nicht die hierfür notwendigen Schlüsselkompetenzen aneignen. Dadurch besteht die Gefahr, dass hier vermehrt Leistungen ins Ausland abfließen.

Potenziale und Use Cases	Besondere Potenziale bestehen, wenn sich EndanwenderInnen aus anderen Bereichen entscheiden, die Technologien im Big Data Stack anzunehmen. Dann ist es möglich, Kompetenzen in IT-Dienstleistungsbetrieben aufzubauen und diese zu Nutzen. Weitere Möglichkeiten bestehen in der Analyse von HörerInnen- und SeherInnenverhalten sowie im besseren Zielgruppenmarketing.

3.2.4.11 Erbringung von Finanz- und Versicherungsdienstleistungen

Finanzdienstleistungen sind im weitesten Sinne alle Dienstleistungen, die einen Bezug zu Finanzgeschäften haben. Diese können sowohl von Kreditinstituten, Finanzdienstleistungsinstituten als auch durch Unternehmen wie Versicherungen, Maklerpools, Bausparkassen, Kreditkartenorganisationen etc. angeboten werden. In Österreich erfasst diese Branche rund 7.000 Unternehmen mit rund 25.000 Beschäftigten. Bekannte Player in diesem Segment sind Unicredit, Erste Bank, Raiffeisen Banken, aber auch Versicherungen wie Generali oder Allianz.

Herausforderungen	Im Bankensektor sind sehr viele Echtzeitanalysen notwendig, welche die Unternehmen vor großen Herausforderungen stellen. Hierbei geht es vor allem darum, Betrugsversuche zu untermauern und zu erkennen. Im Versicherungssektor gilt es vor allem, die jeweiligen Dienstleistungen zu verbessern und anzupassen.
Potenziale und Use Cases	Durch effizientere Algorithmen und Echtzeitdatenanalysen wird es möglich, Betrugsvorgänge besser aufzudecken. Bei Versicherungsprodukten kann Predictive Analytics helfen, diese noch besser für zukünftige Entwicklungen anzupassen. Big Data Analysen bieten auch die Möglichkeit genauere Erkenntnisse über KundInnen zu erlangen in Bezug auf KundInnenbindung und deren Zufriedenheit.

3.2.4.12 Grundstücks- und Wohnungswesen

In dieser Branche sind rund 18.000 Unternehmen mit insgesamt etwa 45.000 Beschäftigten erfasst. Insgesamt wurden zuletzt von diesen Unternehmen Umsatzerlöse in Höhe von 8,5 Mrd. Euro erwirtschaftet. Zu den wichtigsten Unternehmen gehören hier die Immofinanz AG, die Bundesimmobilien GmbH (BIG) oder auch die Buwog - Bauen und Wohnen GmbH

Herausforderungen	Vor allem im städtischen Bereich steigen die Anforderungen an Wohnungen enorm. Das liegt wohl auch darin begründet, dass die Bevölkerung hier stärker als am Lande anwächst. Hier sind vor allem Datenquellen von Interesse, welche Interpretationen auf mögliche Entwicklungen zulassen. Ferner haben die EinwohnerInnen eines Landes auch vielfältige Bedürfnisse, welche individueller adressiert werden können.

Potenziale und Use Cases	Für Großprojekte ergibt es Sinn, detaillierte Analysen über den Markt zu erstellen. Hierbei können demographische Daten sowie Predictive Analytics helfen, herauszufinden, wie sich eine Region oder ein Stadtteil entwickelt. Hierbei entstehen auf Basis von Open Data und Big Data Analysen gerade neue Märkte und Potenziale für echtzeitgetriebene Analysen von zum Beispiel demographischen-geographischen, Karten sowie sozialen Daten für unterschiedliche Anwendungsfälle wie Immobilienvermittlung, Immobilienbewertung sowie Raumplanung.

3.2.4.13 Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen

In dieser Branche sind rund 60.000 Unternehmen mit insgesamt etwa 215.000 Beschäftigten erfasst. Insgesamt wurden zuletzt von diesen Unternehmen Umsatzerlöse in Höhe von 1,4 Mrd. Euro erwirtschaftet.

Herausforderungen	Diese Unternehmenssparte ist vor allem durch EinzelunternehmerInnen geprägt. Hierbei besteht die Herausforderung, dass diese wenig bis gar keine freien Kapazitäten haben, sich um Datenanalysen zu kümmern.
Potenziale und Use Cases	In diesem Sektor besteht vor allem die Möglichkeit, offene Datenquellen in Anwendungen und Dienstleistungen zu integrieren beziehungsweise diese zu neuen Services zu verschmelzen.

3.2.4.14 Öffentliche Verwaltung, Verteidigung, Sozialversicherung

In dieser Branche sind fast 2.600 Unternehmen erfasst. Wichtige Vertreter dieser Branche sind sämtliche Bundesministerien, der Hauptverband der österreichischen Sozialversicherungsträger, Pensionsversicherungsanstalten und Gebietskrankenkassen, etc.

Herausforderungen	Ein zentrales Element der öffentlichen Verwaltung ist die Transparenz, die zunehmend von den jeweiligen Stellen erwartet wird. Dadurch ist es oftmals notwendig, Datenquellen offen zu gestalten, um Vertrauen zwischen den BürgerInnen und der öffentlichen Verwaltung zu schaffen. Vor allem im Sozialversicherungsbereich ist es notwendig, gesundheitsrelevante Trends, Szenarien und Bedrohungen zu finden und darauf zu reagieren. Eine weitere wichtige Herausforderung besteht in der Umsetzung und den rechtlichen Rahmenbedingungen für Datenakquise und Speicherung gerade in Bezug auf das Thema Big Brother. Hierbei müssen in nächster Zeit zukunftsweisende Entscheidungen in Bezug auf die staatliche Haltung zwischen verdachtsunabhängigen Vorratsdatenspeicherungen und den

Freiheitsgraden und Grundrechten der BürgerInnen getroffen werden.

Potenziale und Use Cases	Die öffentliche Verwaltung hat vor allem Potenzial, öffentliche Datenquellen zur Verfügung zu stellen und die Transparenz zu fördern. In den einzelnen Ministerien besteht eine Vielzahl an Möglichkeiten, Analysen zu erstellen, welche Trends erkennen lassen. Im Bereich der Sozialversicherungen ist es vor allem wichtig, Problemfälle im Gesundheitsbereich zu erkennen und auf zukünftige Entwicklungen bereits frühzeitig reagieren zu können. Ein Beispiel sind hier Volkskrankheiten, welche einen enormen volkswirtschaftlichen Schaden anrichten. Durch gezielte Gegenmaßnahmen kann hier gegengesteuert werden. Datenanalysen können hierbei helfen, diese zu erkennen und Maßnahmen zu messen. Soziale Medien bieten Potenziale zur Erkennung von Trends und Bedürfnissen der BürgerInnen in Richtung deren Einbeziehung in politische Prozesse. Big Data Technologien ermöglichen auch die Personalisierung von Dienstleistungen und der damit hergehenden Effizienzsteigerung und Zufriedenheit der BürgerInnen. Potenziale durch Big Data Analysen entstehen auch in der genaueren Vorhersage von finanziellen und budgetären Entwicklung sowie in der Effizienz von Finanzämtern gerade in Bezug auf Steuerbetrug. Auch in den Bereichen Cyber Security und Threat detection können Big Data Technologien erfolgreich eingesetzt werden und zusätzliche Potenziale heben.

3.2.4.15 Erziehung und Unterricht

In dieser Branche sind rund 11.500 Unternehmen erfasst. Diese Branche umfasst nicht nur sämtliche Universitäten und Fachhochschulen Österreichs, sondern auch andere Bildungseinrichtungen bis hin etwa zu Fahrschulen.

Herausforderungen	Die Lehrpläne an Österreichs Bildungseinrichtungen enthalten in vielen Fällen kaum die Informationstechnologie. In einschlägigen Studien an Universitäten und Fachhochschulen gibt es bereits relevante Kurse, es gibt jedoch kein dediziertes Studium für Data Management.
Potenziale und Use Cases	Die Schaffung eines dedizierten Studiums „Data Management“ oder „Data Science“ könnte den Standort Österreich für InvestorInnen interessant machen. Die Grundlage hierfür wäre, dass die Bildungseinrichtungen hochqualifizierte Arbeitskräfte für diesen Themenbereich liefern.

3.2.4.16 Gesundheits- und Sozialwesen

Das Gesundheits- und Sozialwesen umfasst Tätigkeiten der medizinischen Versorgung durch medizinische Fachkräfte in Krankenhäusern und anderen Einrichtungen, der stationären

Pflegeleistungen mit einem gewissen Anteil an medizinischer Versorgung und Tätigkeiten des Sozialwesens ohne Beteiligung medizinischer Fachkräfte. In dieser Branche sind mehr als 44.000 Unternehmen erfasst.

Herausforderungen	Im Krankenbereich gibt es eine Vielzahl an Herausforderungen an datengetriebenen Anwendungen. Besonderes Augenmerk gilt hierbei der Notwendigkeit, bessere Behandlungen zu ermöglichen. Aber auch pharmazeutische Unternehmen haben eine Vielzahl an Herausforderungen im Umgang mit Big Data. Im Sozialwesen stellt sich vor allem die Frage, wie man die ständig alternde Bevölkerung Österreichs besser versorgen kann. Spezifische Herausforderungen ergeben sich in mehreren Bereichen. Die Digitalisierung und die Akquise von Daten, um diese für analytische Problemstellungen verwendbar zu machen, ist ein essenzieller Aspekt für die Bereitstellung zukünftiger innovativer Lösungsansätze. Eine weitere wesentliche Herausforderung besteht in der Integration und Speicherung der Daten. Dieser Bereich bietet enorme Potenziale in der Erforschung von Krankheiten und Behandlungsmethoden. Hierbei müssen rechtliche Rahmenbedingungen unter Berücksichtigung von Datenschutz für die gemeinsame Nutzung und das Teilen von Daten geschaffen werden.
Potenziale und Use Cases	Im Gesundheitswesen wie zum Beispiel bei ÄrztInnen oder in Krankenhäusern können Behandlungen gegebenenfalls zielgerichteter geführt werden, wenn die gesamte Krankengeschichte bekannt ist und ähnliche Fälle verglichen werden können. Hierbei spielen Datenanalysen eine bedeutende Rolle. Medikamente können besser verschrieben werden, wenn die jeweiligen Wechselwirkungen bekannt sind.

3.2.4.17 Kunst, Unterhaltung und Erholung

In dieser Branche sind fast 17.000 Unternehmen mit insgesamt etwa 35.000 Beschäftigten erfasst. Kunst, Unterhaltung und Erholung ist prinzipiell in einem Tourismusland wie Österreich von großer Bedeutung.

Herausforderungen	Die Branche hat eine große Bedeutung im Tourismusland Österreich. Diese Branche ist sehr ähnlich zur Branche Tourismus, hat jedoch keine große IT-Durchdringung. Big Data Technologien zu etablieren ist dadurch wahrscheinlich nicht sonderlich Erfolg versprechend.
Potenziale und Use Cases	Potenziale bestehen vor allem in der KundInnenbindung und der Auffindung von individuellen Bedürfnissen und Eigenschaften einzelner KundInnen. Diese können dann für zielgerichtete Werbung genutzt werden.

3.2.5 Potenzialmatrix

Anhand der Branchendiskussion des vorangegangenen Kapitels wird nun eine Matrix dargestellt, welche abbildet, wie sich die einzelnen Vorteile von datenbezogenen Anwendungen in den einzelnen Branchen auswirken. Hierbei wird in groß, mittel und gering unterschieden. Die drei Kategorien sind:

- Generelle IT-Leistungsfähigkeit der Branche. Hier wird diskutiert, wie die aktuelle Bedeutung der IT in der Branche ist. Manche Branchen haben bereits eine sehr hohe IT-Durchdringung, während diese in anderen Branchen noch sehr niedrig ist.
- Potenzial der Branche, Big Data Anwendungen einzuführen. Hier wird diskutiert, ob in der Branche aktuell Big Data Anwendungen diskutiert werden und ob IT-EntscheiderInnen der Branche dies ernsthaft überlegen.
- Potenzial der Branche, durch Big Data Wettbewerbsvorteile zu gewinnen. Hier wird analysiert, ob Unternehmen dieser Branchen Wettbewerbsvorteile gewinnen können, wenn Big Data eingeführt wird. Wenn das Potenzial hierbei als gering oder mittel eingestuft wird, so kann dies bedeuten das hier bereits ein hoher Wettbewerbskampf mithilfe von datengetriebenen Anwendungen stattfindet.

Branche	Generelle IT-Leistungsfähigkeit der Branche	Potenzial der Branche, Big Data Anwendungen einzuführen	Potenzial der Branche, durch Big Data Wettbewerbsvorteile zu gewinnen
LAND- UND FORSTWIRTSCHAFT, FISCHEREI			
BERGBAU UND GEWINNUNG VON STEINEN UND ERDEN			
HERSTELLUNG VON WAREN			
ENERGIEVERSORGUNG			
WASSERVERSORGUNG; ABWASSER- UND ABFALLENTSORGUNG UND BESEITIGUNG VON UMWELTVERSCHMUTZUNGEN			
BAU			
HANDEL; INSTANDHALTUNG UND REPARATUR VON KRAFTFAHRZEUGEN			
VERKEHR UND LAGEREI			
BEHERBERGUNG UND GASTRONOMIE			
INFORMATION UND KOMMUNIKATION			
ERBRINGUNG VON FINANZ- UND VERSICHERUNGSDIENSTLEISTUNGEN			
GRUNDSTÜCKS- UND WOHNUNGSWESEN			

ERBRINGUNG VON FREIBERUFLICHEN, WISSENSCHAFTLICHEN UND TECHNISCHEN DIENSTLEISTUNGEN			
ÖFFENTLICHE VERWALTUNG, VERTEIDIGUNG; SOZIALVERSICHERUNG			
ERZIEHUNG UND UNTERRICHT			
GESUNDHEITS- UND SOZIALWESEN			
KUNST, UNTERHALTUNG UND ERHOLUNG			

Tabelle 7: Branchenbedeutung

Legende

Groß
Mittel
Gering

3.2.6 Domänenübergreifende Anforderungen und Potenziale

Im Bereich Big Data entstehen unabhängig von der Branche und dem Anwendungsfeld unterschiedliche Herausforderungen. Hierbei wird unabhängig von der jeweiligen Branche auf Herausforderungen und Potenziale von Big Data in bestimmten Abteilungen eines Unternehmens eingegangen. Die Anwendungsfälle sind hierbei über die Industrien gelegt, was eine Anwendung in den meisten in Kapitel 3.2.4 gelisteten Industrien möglich macht. Die Querschnittsthemen fokussieren auf die Verwaltung eines Unternehmens.

3.2.6.1 Marketing und Sales

Marketing und Sales ist eine wichtige Querschnittsfunktion in sämtlichen Unternehmen. Marketing ist hierbei sehr wichtig für Unternehmen, frei nach dem Zitat „wer nicht wirbt, der stirbt“ von Henry Ford, dem Gründer des gleichnamigen US-Automobilkonzerns. Hierbei gibt es jedoch starke Schwankungen von Sektor zu Sektor. Unternehmen, welche im B2B Bereich operieren, benötigen oftmals ganz andere Konzepte als jene, die sich direkt an die Endkunden wenden. So braucht ein Stahlkonzern oder ein Konzern, welcher Maschinenbauteile herstellt, oftmals gar kein effektives Marketing, sondern nur einen guten Verkauf, welcher sich an andere Unternehmen wendet.

Herausforderungen

In einigen Branchen ist der Wettbewerbsdruck sehr hoch geworden. Dies äußert sich durch oftmalige Sparmaßnahmen. Für diese Unternehmen ist es notwendig, sich durch Marketing und Sales vom Mitbewerber abzuheben. Eine wesentliche Herausforderung besteht jedoch in der Zugänglichkeit von Marktdaten, welche oftmals teuer zugekauft werden müssen, sowie der rechtlichen Verwendbarkeit der Daten. In vielen Fällen sind diese personenbezogen, was mehrere Implikationen in Bezug auf die Rechtslage und Datenschutz nach sich zieht.

Potenziale und Use Cases	Vor allem durch die Kombination von offenen Datenquellen und der Überschneidung von vorhandenen Datenquellen des Unternehmens ergeben sich für das Unternehmen enorme Potenziale. Dies kann sogar so weit gehen, dass diverse Daten wieder zurück in offene Datenquellen fließen. Eine bessere Analyse von Daten bringt den Unternehmen entscheidende Verbesserungen im Marketing, welches auch den Standort Österreich zugutekommen kann. Um diesen Themenkreis herum kann sich ferner eine lebhafte Industrie entwickeln, wodurch relevante Marketingprodukte entstehen können, die internationale Bedeutung erlangen können.

3.2.6.2 Finanzen und Controlling

Die Abteilungen Finanzen und Controlling sind in fast jedem Unternehmen meist sehr einflussreiche Abteilungen. Oftmals ist es so, dass der CFO (Vorstand der Finanzabteilung) auch der zukünftige CEO ist. Sämtliche Unternehmen sind finanzgetrieben, eine effiziente Finanz bietet dem Unternehmen vielfältige Möglichkeiten.

Herausforderungen	Vor allem durch die Krise seit 2009 und der daraus langsamen Erholung der Märkte sind die Unternehmen nach wie vor unter einem starken Finanzdruck. Auf der anderen Seite flüchten immer mehr Anleger vom klassischen Sparbuch in Aktien, da diese eine bessere Rendite versprechen. Zentrale Herausforderungen in den Finanzabteilungen sind oftmals auf die unterschiedlichen Industrien zurückzuführen. So ist es bei produzierenden Unternehmen wesentlich, eine ideale Preisfindung zu ermöglichen, um den Gewinn ideal anzupassen. Komplexe Probleme ergeben sich auch mit Prognosen, welche oft für Monate oder Jahre im Vorhinein gegeben werden sollen.
Potenziale und Use Cases	Durch die Integration von Big Data Anwendungen kann man viele der oben genannten Herausforderungen adressieren. Durch die Integration von demographischen Daten kann man Modelle errechnen, welche die Preiselastizität analysieren. Hierbei ergeben sich Querschnitte mit dem Marketing. Besonders die Gewinn- und Umsatzprognose ist ein schwieriges Thema. Dieses kann mit Big Data und Predictive Analytics sehr erfolgreich adressiert werden.

3.2.6.3 Personalwesen (HR)

Das Personalwesen im Unternehmen hat sich vor allem in großen und mittleren Unternehmen von einer reinen „Einstellabteilung“ hin zu einer Abteilung für die MitarbeiterInnenentwicklung weiterentwickelt. Hierbei geht es vorrangig darum, qualifiziertes Personal zu halten, und weiter zu qualifizieren. Die Devise lautet oftmals, dass MitarbeiterInnen als Kapital und nicht als Kostenstelle geführt werden sollen.

Herausforderungen	Vor allem der Fachkräftemangel erschwert es den Unternehmen,

geeignetes Personal zu finden. Oftmals muss hierbei das Personalwesen lenkend eingreifen und einen Ausbildungspfad entwickeln. Qualifiziertes Personal lässt sich meist nicht leicht im Unternehmen halten, was für das Unternehmen dann einen erheblichen Verlust darstellt. Dadurch sind ebenfalls Maßnahmen zur Karriereentwicklung notwendig. In modernen Unternehmen besteht ferner die Herausforderung, die ideale Person für das Team zu finden. Dies ist oftmals nicht einfach zu erreichen.

Potenziale und Use Cases: Durch die Integration von Datenanalysen und zukünftigen Szenarien wird es möglich, Entwicklungen besser abschätzen zu können und darauf aufbauend einen Ausbildungs- und Qualifizierungspfad zu erstellen. Mithilfe von Analysen von BewerberInnen kann es möglich werden, die zukünftige Teamzusammensetzung besser abschätzbar zu machen. Da es sich hierbei jedoch um personenbezogene Daten handelt, gilt es dieses stark zu hinterfragen.

3.2.6.4 Einkauf

Einkaufsabteilungen sind an sich in jeglichen Unternehmen vorhanden. Hierbei kann es jedoch von Branche zu Branche wesentliche Unterschiede geben. Wesentlich zu unterscheiden ist vor allem die Lagerfähigkeit und -notwendigkeit.

Herausforderungen: Vor allem in Branchen, wo Waren stark rollieren (Einzel- und Großhandel) sowie in Branchen mit zeitkritischer Produktion (nach dem Just-In-Time-Prinzip) bestehen große Herausforderungen. Wesentlich ist hierbei die Frage, wie das Produktsegment optimiert werden kann und welche Produkte wann zur Verfügung stehen sollen.

Potenziale und Use Cases: Durch die Analyse von demographischen Daten aus offenen Quellen und der zu erwartenden langfristigen Entwicklung kann das Produktsegment vor allen im Handel stark verbessert werden. Um die Lagerkosten zu reduzieren, bedarf es einer verstärkten Analyse der zu erwartenden Auslastung, welche auch auf einzelne Tage abgebildet werden kann.

3.2.6.5 Forschung und Entwicklung

In der Forschung und Entwicklung sind verteilte Systeme, worunter auch Big Data fällt, schon sehr lange relevant. Hierzu zählt vor allem die Ausführung von langlaufenden Algorithmen, wobei hierfür viele Daten produziert und auch analysiert werden.

Herausforderungen: Die Forschung und Entwicklung ist oftmals mit geringen Budgets konfrontiert. Auf der anderen Seite sind jedoch große IT-Leistungen notwendig. Die Forschung und Entwicklung hat oft die Notwendigkeit, das Unternehmen mit neuen Produkten zu versorgen und auch zu unterstützen.

Potenziale und Use Cases	Potenzial besteht vor allem durch die Ausführung von Datenanalysen in der Cloud. Hierdurch besteht wesentliches Einsparungspotenzial. Ferner werden durch eine Cloud-basierte Ausführung Ressourcen frei – man muss sich nicht tagelang mit dem Setup beschäftigen, sondern kann dieses durch wenige Mausklicks durch Cloud-Plattformen erstellen.

3.2.6.6 Qualitätssicherung

Vor allem in der produzierenden Industrie aber auch in einigen weiteren Domänen ist die Qualitätssicherung wesentlich für den Unternehmenserfolg. Hierbei geht es darum, die Kundenzufriedenheit zu heben und die durch Fehlproduktionen entstehenden Schäden zu minimieren.

Herausforderungen	Material- und Produktionsfehler sind oft sehr teuer für das Unternehmen. Um diese Kosten zu reduzieren, ist es notwendig, eine erfolgreiche Qualitätssicherung zu erstellen. Da durch diese Arten von Fehlern große finanzielle Schäden entstehen, ist hier ein besonderer Leistungsdruck zu erwarten.
Potenziale und Use Cases	Durch den verstärkten Einsatz von Datenanalysen, welche auch in Echtzeit erfolgen, können Fehler reduziert werden. Innovative Industrielösungen können auf dieser Basis erstellt werden. Hierdurch kann vor allem Österreich seine Wettbewerbsfähigkeit stärken und eine USP erreichen.

3.2.6.7 KundInnenservice

Das KundInnenservice trägt verstärkt zur Zufriedenheit der KäuferInnen bei. Kann man sich hier etablieren, so werden die KundInnen immer wieder zurückkehren und man sorgt damit für eine höhere KundInnenbindung. Dies bietet vor allem in stark umkämpften Märkten und Branchen starke Vorteile.

Herausforderungen	Der Druck, KundInnen langfristig an das Unternehmen zu binden, ist vor allem im Handel wesentlich. Hierbei können jene Unternehmen, die einen guten KundInnenservice anbieten, Vorteile erzielen. Herausforderungen bestehen darin, aus bestehenden Problemen zu lernen und diese in Wissen für das Unternehmen umzuleiten.
Potenziale und Use Cases	Vor allem durch die Analyse der KundInnenservice-Vorfälle kann das Unternehmen wesentliche Elemente über seine KundInnen lernen und somit das Service ständig verbessern. Werden diese Daten mit weiteren Daten angereichert, kann die Wettbewerbssituation bedeutend gestärkt werden.

3.2.7 Domänenübergreifende Anforderungen in der Datenverarbeitung

In diesem Abschnitt werden generelle technologische Anforderungen nach spezifischen durchzuführenden Schritten in der Umsetzung von Big Data beschrieben und diskutiert. Hierbei wird auf Anforderungen in den Bereichen Datenakquise, Datenspeicherung, Data Curation und der Verwendung von Daten fokussiert. Diese wurden auf Basis von Interviews und des „Consolidated Technical White Papers“ des Projekts Big Data Public Private Forum (Big Data Public Private Forum, 2013) ausgearbeitet.

3.2.7.1 Datenakquise

Die Datenakquise beschäftigt sich mit der Erfassung, der Filterung und dem Aufbereiten der Daten. Die größte Herausforderung in der Akquise von Daten besteht in der Bereitstellung von Frameworks welche die aktuellen riesigen Datenquellen in Echtzeit bearbeiten können (z.B. Sensordaten) und somit die Filterung und die Aufbereitung der Daten für deren Weiterverarbeitung ebenfalls in Echtzeit durchführen können. Die wichtigsten Anforderungen in diesem Bereich sind wie folgt:

- Vielfalt an Datenquellen: Eine wichtige technische Anforderung im Bereich Datenakquise ist die Möglichkeit unterschiedliche Datenquellen in verschiedenen Datenformaten adressieren zu können.
- Bereitstellen von Schnittstellen: Neben der Unterstützung von diversen Datenformaten soll ein Framework zur Akquise von Daten auch die einfache Anbindung an unterschiedliche Speicherlösungen sowie Analysetools bieten. Hierfür werden generische programmiersprachenunabhängige APIs benötigt.
- Effektive Vorverarbeitung von Daten: Ein Framework für die Akquise von Daten muss für die jeweiligen aus den Datenquellen entstehenden Herausforderungen gewappnet sein. Hierbei ist wichtig, dass zum Beispiel unstrukturierte Daten in strukturierte Formate umgewandelt werden können, um deren Speicherung und Weiterverarbeitung zu vereinfachen.
- Bild, Video und Musikdaten: Eine große Herausforderung entsteht in der Datenakquise von Mediendaten. Hierbei entstehen neue technologische Herausforderungen, welche von derzeitigen Big Data Frameworks noch nicht gesondert behandelt werden.

3.2.7.2 Datenspeicherung

Datenspeicherung beschäftigt sich mit der Speicherung, der Organisation und der Verwaltung von großen Datenmengen und kann technologisch in NoSQL Systeme und verteilte Dateisysteme unterteilt werden (siehe Kapitel 2.2.4.1). Die größten Herausforderungen und Anforderungen in diesem Bereich können wie folgt zusammengefasst werden:

- Skalierbare Hardwarelösungen: In der Bereitstellung und der Entwicklung von neuen Hardwarelösungen für die Speicherung von riesigen Datenmengen sind in letzten Jahren große Fortschritte erzielt worden. Die Speicherung der Datenmengen ist mittlerweile leistbar und nicht mehr das entscheidende Kriterium. Nichtsdestotrotz stellen sich hier große Herausforderungen in der Auswahl der Hardwarelösungen für spezifische Problemstellungen.
- Datenorganisation und Modellierung: Im Big Data Bereich geht der Trend zu der Speicherung der Rohdaten und der nachträglichen Modellierung und Extraktion von Wissen. Im Unterschied zu relationalen Datenbanken ergibt sich hierbei die große Herausforderung der Modellierung der Daten und der Integration der Daten mit anderen Datenquellen in am besten interaktiver aber automatisierter Art und Weise. Hierbei entstehen neue Anforderungen an Big Data Speicherlösungen in Richtung Datenmodellierung,

Datendefinitionssprachen, Schemadefinitionen und Echtzeitumlegung und Integration von unterschiedlichen Daten.

- CAP-Theorem: Wichtige Herausforderungen im Bereich von Big Data Speichersystemen ergeben sich in der Umsetzung von Systemen in Bezug auf das CAP-Theorem. Derzeit bieten alle vorhandenen Systeme Schwächen in einem Bereich (Konsistenz, Partitionstoleranz, Verfügbarkeit). Hierbei ist ein Trend der Zusammenführung von Funktionalitäten relationaler Datenbanken und NoSQL Systemen zu beobachten der neue technologische Herausforderungen birgt. Domänenunabhängig stellt sich derzeit hierbei die Abwägung der projektspezifischen Anforderungen in Bezug auf diese Charakteristiken.
- Kompression, Sicherheit und Verschlüsselung: Wichtige Herausforderungen und Anforderungen ergeben sich in der Bereitstellung von Systemen welche Verschlüsselung der Daten und sicheren Zugriff auf diese gewährleisten. Derzeit sind Big Data Speicherlösungen in diesen Aspekten meistens noch nicht auf dem Stand von relationalen Datenbanken. Die transparente Unterstützung von Komprimierungsmethoden ist ebenfalls eine wesentliche Herausforderung in Bezug auf die benötigte Speicherkapazität.

3.2.7.3 Data Curation

Data Curation beschäftigt sich mit der aktiven und durchgängigen Verwaltung von Daten in Bezug auf deren gesamten Lebenszyklus. Wichtige Punkte hierbei sind die Aufrechterhaltung der Datenqualität, die Ermöglichung der Wiederverwendung von Daten und auch die Bewahrung der Daten über längere Zeiträume. Die größten Anforderungen in diesem Bereich können wie folgt zusammengefasst werden:

- Skalierende Methoden für Data Curation: Die aktuell verfügbaren Methoden für die Bestandserhaltung von Daten müssen an neue Gegebenheiten durch Big Data angepasst werden. Um diese Methoden auf große und vielfältige Daten anwenden zu können wird in Zukunft ein großes Maß an Automatisierung benötigt werden.
- Neue Datentypen: Derzeitige Ansätze für die Bestandserhaltung von Daten beschäftigen sich meistens mit strukturierten und semi-strukturierten Daten. Diese müssen in weiterer Folge auf alle Arten von Datenquellen ausgebaut werden was unter anderem Multimediadaten und Sensordaten miteinschließt.
- Integrativer Zugang: Um einen ganzheitlichen Bestandserhaltungsansatz zu gewährleisten wird Expertenwissen aus unterschiedlichen Domänen benötigt. Eine Herausforderung besteht hierbei in der Umsetzung von integrativen Ansätzen welche Domänenwissen, Big Data Experten sowie Data Curation Experten miteinbezieht.

3.2.7.4 Verwendung von Daten

Der Bereich Verwendung von Daten spiegelt die Ebene Utilization des Big Data Stacks (siehe Kapitel 2.2.1) wider und beschäftigt sich mit der Generierung von Mehrwert aus Daten. Zukünftige Anforderungen in diesem Bereich entstehen vor allem in den folgenden Aspekten:

- Vertrauen in Daten: Um große Datenmengen für Entscheidungen heranziehen zu können ist das Vertrauen in die vorhandenen Datenquellen enorm wichtig. Hierfür ist einerseits die Qualität der Daten ein entscheidender Aspekt. Diese sollten mit Metadaten gemeinsam zur Verfügung stehen, um auf diesen Daten basierende Entscheidungen auch im Nachhinein verifizieren zu können und diese nachvollziehbar zu machen.
- Echtzeitzugriff: Derzeit verfolgen die meisten Big Data Ansätze eher den Batch Zugriff beziehungsweise werden „Near-Real-Time“-Analysen der Daten zur Verfügung gestellt. Eine

wichtige Anforderung hierbei ist die Bereitstellung von Echtzeitanalysen auf Basis der Rohdaten. Dieser Aspekt beinhaltet viele technologische Herausforderungen sowie die Integration von Tools auf den unterschiedlichen Ebenen des Big Data Stacks.

- Strategische Entscheidungen: Derzeit werden die meisten Datenanalysen auf Basis konkreter Fragestellungen erstellt und anschließend durchgeführt. Ein wesentlicher Aspekt der Zielsetzungen im Bereich Big Data ist es aufgrund der zur Verfügung stehenden Daten neue innovative Fragestellungen auf einfache Art und Weise zu unterstützen. Die Grundlage hierfür ist durch das Vorhandensein der Daten und der Möglichkeit der einfachen Integration dieser gegeben. Die Umsetzung von Echtzeitanalysen für neue Fragestellungen wird hier einen zukünftigen Meilenstein in Big Data Technologiebereich bilden.
- Expertise: Für die effiziente Verwendung von Daten ist Kompetenz in der Interpretation und Aufbereitung der Daten notwendig welche derzeit teilweise am Arbeitsmarkt nicht verfügbar ist. Eine detaillierte Analyse dieser Anforderung wird in Kapitel 4.2.1.2 durchgeführt.

4 Best Practice für Big Data-Projekte

Der effiziente Einsatz von Big Data Technologien kann MarktteilnehmerInnen zusätzliches Wissen liefern, aber auch neue innovative Dienste generieren, die einen erheblichen Mehrwert mit sich bringen. Andererseits erfordert der effiziente Einsatz von Big Data Technologien fachspezifisches Wissen in unterschiedlichen Bereichen, beginnend bei den zu verwendenden Technologien, über Wissensextraktion und Aufbereitung, bis hin zu rechtlichen Aspekten. Diese breite thematische Aufstellung des Themas Big Data, von den rein technischen Aspekten für die Datenspeicherung und Verarbeitung, dem Wissensmanagement, den rechtlichen Aspekten für die Datenverwendung sowie den daraus resultierenden Potenzialen für Unternehmen (neue Dienste und Produkte sowie neues Wissen generieren) erfordert umfassende Kompetenzen für deren effiziente Umsetzung in spezifischen Projekten und für die Entwicklung neuer Geschäftsmodelle.

Dieses Kapitel bietet Organisationen einen Leitfaden für die Umsetzung und Anwendung von Big Data Technologien um das Ziel der erfolgreichen Abwicklung und Implementierung von Big Data in Organisationen zu erreichen. Es werden Best Practices Modelle anhand von spezifischen Big Data Leitprojekte im Detail erörtert und anschließend wird auf dessen Basis ein spezifischer Leitfaden für die Umsetzung von Big Data Projekten in Organisationen präsentiert. In diesem Rahmen wird ein Big Data Reifegrad Modell, ein Vorgehensmodell, eine Kompetenzanalyse, unter anderem des Berufsbilds Data Scientist, sowie eine Referenzarchitektur für die effiziente Umsetzung von Big Data vorgestellt.

4.1 Identifikation und Analyse von Big Data Leitprojekten

Big Data umfasst unterschiedliche Technologieebenen und Charakteristiken. Des Weiteren werden Big Data Technologien in vielen verschiedenen Domänen eingesetzt und es ergeben sich jeweils differenzierte Anforderungen und Potenziale. In Kapitel 2 wurde der Bereich Big Data definiert und es wurden Technologien sowie Marktteilnehmer in Österreich erhoben. In Kapitel 3 wurden Domänen identifiziert in welchen Big Data als relevante Technologie schon Einzug gehalten hat und zukünftige Potenziale gesehen werden. Diese domänenspezifischen Anforderungen und Potenziale sind gesondert herausgearbeitet. In vielen der oben analysierten Domänen wurden oder werden Big Data Projekte durchgeführt welche großen Einfluss auf die bereitgestellten Technologien, aber vor allem auf die jeweilige Domäne haben können. In diesem Kapitel werden mehrere Projekte aus unter anderem aus den Bereichen Mobilität und Gesundheitswesen vorgestellt und im Detail auf Big Data relevante Aspekte analysiert. Die vorgestellten Projekte bilden viele der essenziellen Technologien ab und haben großes Potenzial, den österreichischen Markt in Bezug auf Big Data weiterzuentwickeln. Des Weiteren wurden im Rahmen der Studie viele zusätzliche Projekte aus den unterschiedlichsten Bereichen im Rahmen von Interviews erhoben und diskutiert. Die Erkenntnisse aus diesen Erhebungen sind in der Umsetzung des Leitfadens sowie der Markt und Potenzialanalyse eingearbeitet.

Hier finden sie eine Kurzübersicht über die nachfolgend im Detail beschriebenen Projekte:

Projektname	Domäne	Projekttyp	Leitung (Österreich)	Abstract
Real-Time Data Analytics	Mobility	Internes Projekt	Austrian Institute of Technology	Bereitstellung einer flexiblen Infrastruktur für die Echtzeitanalyse von großen und dynamischen Datenströmen für mobilitätsbezogene Forschung
VPH-Share	Medizin	EU Projekt	Universität Wien	VPH-Share entwickelt die Infrastruktur und Services für (1) die Bereitstellung und gemeinsame Benutzung von Daten und Wissen, (2) die Entwicklung von neuen Modellen für die Komposition von Workflows, (3) die Förderung von Zusammenarbeit unterschiedlicher Stakeholder in dem Bereich Virtual Physiological Human (VPH).
Leitprojekt Handel	Handel	Nationales Projekt		Echtzeitdatenanalyse in Filialen
Katastrophenerkennung mit TRIDENC	Industrie	Nationales Projekt	Joanneum Research	Konzeption und Entwicklung einer offenen Plattform für interoperable Dienste, die ein intelligentes, ereignisgesteuertes Management von sehr großen Datenmengen und vielfältigen Informationsflüssen in Krisensituationen
Prepare4EODC	Weltraum	Nationales Projekt	TU Wien	Förderung der Nutzung von Erdbeobachtungsdaten für die Beobachtung globaler Wasserressourcen durch eine enge Zusammenarbeit mit Partnern aus der Wissenschaft, der öffentlichen Hand als auch der Privatwirtschaft.

Tabelle 8: Überblick der analysierten Projekte

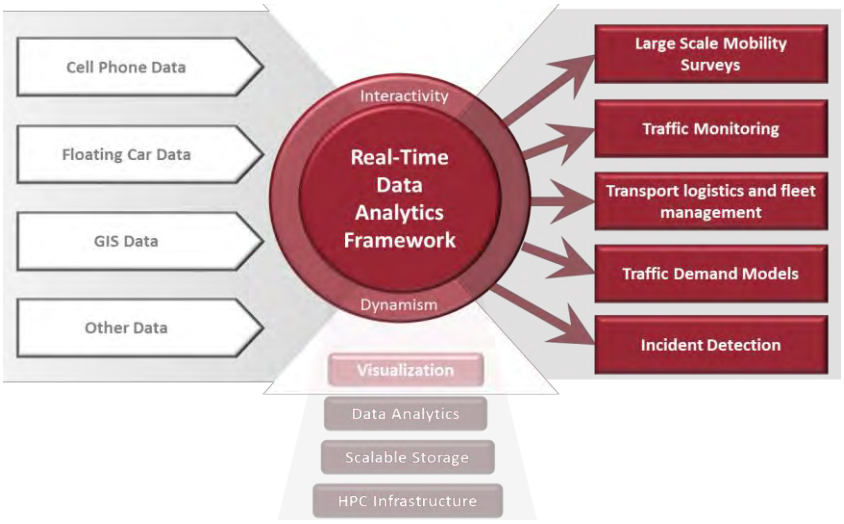
Nachfolgend werden diese Projekte näher beschrieben und der jeweilige Lösungsansatz wird detailliert präsentiert. Des Weiteren werden die Ergebnisse der Projektanalysen in Bezug auf die Big Data Charakteristiken (Volume, Veracity, Velocity, Value) sowie auf die verwendeten Technologien in Bezug auf den Big Data Stack (Utilization, Analytics, Platform, Management) vorgestellt. Für jedes Projekt wurden die jeweiligen projektspezifischen Anforderungen und Potenziale erhoben welche im Detail aufgeführt und analysiert sind.

4.1.1 Leitprojekt Verkehr: Real-time Data Analytics for the Mobility Domain

Real-time Data Analytics for the Mobility Domain	
Domäne	Transport
Projekttyp	Internes Projekt
Projektpartner	Austrian Institute of Technology (AIT) Mobility Department
Ansprechpartner	DI (FH) Markus Ray, AIT Tel: +43 50550 6658 Email: markus.ray@ait.ac.at
Projektbeschreibung	Dieses Projekt beschäftigt sich mit der Bereitstellung einer flexiblen Infrastruktur für die Echtzeitanalyse großer und dynamischer Datenströme für mobilitätsbezogene Forschung. Die Herausforderungen bestehen in der Verarbeitung und Verschneidung großer Datenmengen mit Mobilitätsbezug und deren Bereitstellung in Echtzeit. In den letzten Jahren ist die verfügbare Menge an für diese Domäne relevanten Daten enorm angestiegen (z.B. Verkehrssensoren wie Loop Detectors oder Floating Car Daten, Wetterdaten) und neue Datenquellen (z.B. Mobilfunkdaten oder Social Media Daten) stehen für detaillierte Analysen zur Verfügung. Dies schafft enorme Potenziale für innovative Lösungsansätze, die beispielsweise dynamische Aspekte (wie Echtzeitverkehrszustand) bei multi-modalen Tourenplanungen berücksichtigen (Bsp. Krankentransport), die semi-automatisierte Durchführung von vertieften Mobilitätserhebungen ermöglichen (z.B. für Verkehrsnachfrageerfassung), neuartige Erfassung multi-modaler Verkehrszustände etablieren oder die Bereitstellung profilbasierter individueller Mobilitätsinformationen ermöglichen. Solche Anwendungen erfordern technologisch innovative Konzepte, die diese Datenmengen in Echtzeit verarbeiten und analysieren können.
Lösungsansatz	Die technologische Lösung basiert auf der flexiblen Integration von High-Performance Computing (HPC) und Big Data Lösungen. Dieses integrierte Analyseframework nimmt folgende Herausforderungen in Angriff: • Effiziente verteilte Speicherung und Abfrage großer Datenmengen (inkl. anwendungsbezogener in-memory

Lösungsansätze).

- Schnelle Laufzeit durch den Einsatz von HPC Konzepten
- Flexible modul-basierte Definition komplexer Analysearbeitsabläufe
- Unterstützung der Integration existierender Anwendungen
- Echtzeitverarbeitung und Integration heterogener Datenstreams



Eine integrierte modulare Data Pipeline mit dem MapReduce Programmierparadigma unterstützt die flexible Kombination unterschiedlicher applikationsspezifischer Module zu komplexen Arbeitsabläufen. Einerseits werden entsprechende Module für den Import und Export spezifischer Datenquellen angeboten, andererseits sind unterschiedliche Machine Learning Algorithmen und Funktionen implementiert. Diese können flexibel in der Map- sowie in der Reduce-Phase kombiniert werden, oder gesondert auf den HPC Ressourcen ausgeführt werden. Dies ermöglicht die flexible Ausnutzung unterschiedlicher Hardwareressourcen und Speichersysteme für applikationsspezifische Problemstellungen mit dem Ziel einer in Echtzeit durchgeführten Datenanalyse.

Anforderungen

Applikationsspezifische Anforderungen:

- Optimierung der multi-modalen Verkehrssystems: Bspw. Erhebung und Modellierung der Verkehrsnachfrage, Verkehrsüberwachung und Steuerung, Transportlogistik und Flottenmanagement, Empfehlen und Überprüfen von Verkehrsmaßnahmen.

Soziale Anforderungen:

- Richtlinien für den Umgang mit persönlichen Mobilitätsdaten

Rechtliche Anforderungen:

- Privacy und Security: Klare Regelung in Bezug auf personenbezogene Daten (rechtliche Rahmenbedingungen, Datenschutz, Sicherheit)

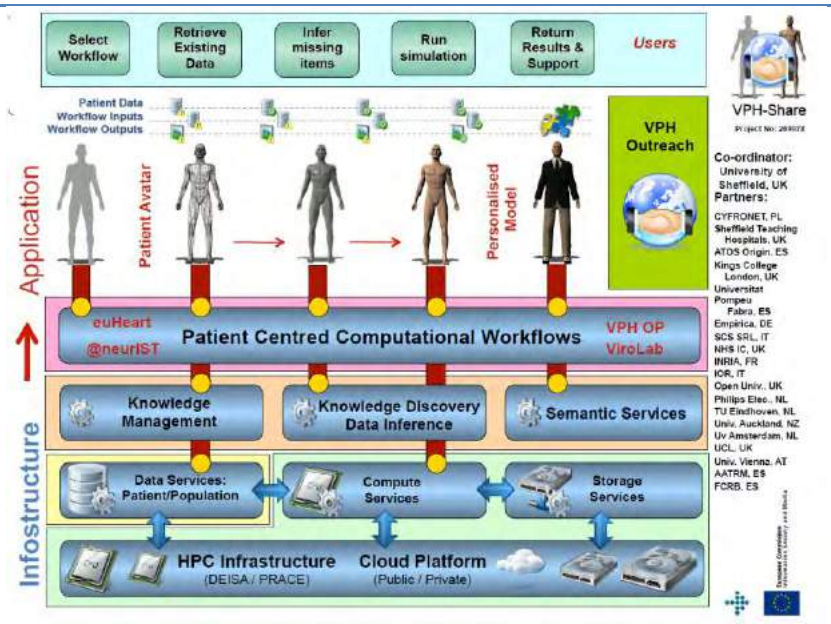
Big Data Charakteristiken *(Welche Big Data Charakteristiken werden adressiert und in welchem Umfang?)*	**Volume**	Sehr große Datenmengen in den Bereichen Mobilfunkdaten, Floating Car Data und Social Media
	Variety	Komplexität in Bezug auf unterschiedlichste Datenstrukturen und -formate von Binärdaten, GIS-Daten bis zu Freitext
	Veracity	Komplexität in Bezug auf Echtzeitanbindung und Integration von unterschiedlichen Datenstreams
		Komplexität in Bezug auf Echtzeitausführung von komplexen Analysealgorithmen auf Basis von großen Datenmengen
	Value	Integration und Echtzeitanalyse von unterschiedlichen Datenquellen birgt enormes Potenzial in innovativen Methoden für unterschiedlichste Mobilitätsanwendungen
Big Data Stack *(Welche Big Data Ebenen und welche Technologien werden adressiert und verwendet?)*	**Utilization**	Die Analyse Ergebnisse können flexibel in Web-basierte Oberflächen integriert werden. Ziel ist die interaktive Analyse integrierter Datenquellen.
	Analytics	Das Analyseframework bietet Implementierungen diverser Machine Learning Algorithmen. Diese können auf Basis der darunterliegenden Plattform sowohl auf HPC Infrastruktur als auch Daten-lokal in Datenzentren ausgeführt werden.
	Platform	Das Projekt bedient sich des Hadoop Ökosystems und bindet mehrere Plattformen für die Ausführung von Daten-intensiven Applikation ein. Die Hadoop Ausführungsumgebung ist nahtlos in eigene Entwicklungen eines Pipelining Frameworks integriert und ermöglicht auf diese Weise, die skalierbare Analyse von (binären) Daten auf Basis applikationsspezifischer Algorithmen. Die enge Verzahnung des MapReduce Paradigmas mit Pipelining Konzepten ermöglicht die Daten-lokale Ausführung komplexer Analysealgorithmen.
	Management	Innerhalb des Projekts werden unterschiedliche Datenspeicherungssysteme aus dem Hadoop Ökosystem verwendet. Die Storage und Rechenressourcen sind auf HPC Ressourcen und lokale Big

	Data Cluster aufgeteilt.
Potenziale	Bereitstellung eines Frameworks auf Basis von HPC und Big Data Technologien das die flexible Integration von Echtzeitdaten und zusätzlicher Datenquellen für innovative Forschungsmethoden im Mobilitätsbereich ermöglicht - Die gemeinschaftliche und integrierte Bereitstellung verschiedener Datenquellen fördert die Entwicklung neuer Forschungsfragen und innovativer Geschäftsfälle in den Bereichen HPC, Big Data sowie Mobilität. - Das erstmalige Ermöglichen eines gemeinsamen Zugriffs auf verschiedene verkehrsbezogene Datenquellen ermöglicht die Entdeckung neuer und die Verbesserung bestehender Verkehrsmodelle. - Die innovative Kombination von HPC und Big Data Technologien für komplexe Echtzeitdatenanalysen dient als Vorzeigebeispiel und liefert Anstöße für weitere Projekte.

4.1.2 Leitprojekt Healthcare: VPH-Share

Projekt „VPH-Share: The Virtual Physiological Human, Sharing for Healthcare - A Research Environment“	
Domäne	Healthcare
Projekttyp	EU Projekt mit österreichischer Beteiligung (ICT for Health – Resource book of eHealth Projects - FP7)
Projektpartner	University of Sheffield, Coordinator Cyfronet Sheffield Teaching Hospitals Atos Kings College London Unversity of Pompeu Fabra, Spain Empirica SCS Supercomputing Solutions INRIA Instituto Ortopedico Rizzoli di Bologna Phillips The Open University Technische Universiteit Eindhoven The University of Auckland University of Amsterdam UCL **Universität Wien**

	Agència de Qualitat i Avaluació Sanitàries de Catalunya Fundació Clínic Barcelona
Ansprechpartner	Norman Powell BSc PhD VPH-Share Project Manager University of Sheffield Univ.-Prof. Dipl.-Ing. Dr. Siegfried Benkner Universität Wien Email: siegfried.benkner@univie.ac.at
Projektbeschreibung	Das Projekt VPH-Share ermöglicht die Integration von sorgfältig optimierten Services für die gemeinsame Benutzung und Integration von Daten, die Entwicklung von Modellen und die gemeinschaftliche Zusammenarbeit von unterschiedlichen Stakeholdern im Gesundheitsbereich auf Basis einer europäischen Cloud Infrastruktur. Innerhalb von VPH-Share werden meistens klinische Daten von individuellen Patienten (medizinische Bilder, biomedizinische Signale, Bevölkerungsdaten, ...) verarbeitet. Das Projekt umfasst unterschiedlichste Varianten von Operationen in Bezug auf diese Daten. Diese umfassen die sichere Speicherung, den sicheren Zugriff, die Annotierung, Inferenz und Assimilation, komplexe Bildverarbeitungsmechanismen, mathematische Modellierungen, die Reduktion der Daten, und die Repräsentation von aus den Daten generiertem Wissen. Das Projekt fokussiert auf die größte Herausforderung in diesem Bereich: der Schnittstelle zwischen dem Reichtum an Daten von medizinischen Forschungseinrichtungen und klinischen Prozessen. Das Ziel der Schaffung einer umfassenden Infrastruktur für die gemeinschaftliche Bereitstellung von Daten, Informationen und Wissen, deren Verarbeitung an Hand von Workflows und deren Visualisierung wird von einem europäischen Konsortium mit einem breiten Kompetenzmix in Angriff genommen und an Hand von vier Flagship Workflows (existierende FP6 und FP7 Projekte @neurIST, euHeart, VPHOP, Virolab) evaluiert.
Lösungsansatz	Das VPH-Share Projekt basiert auf einer Infrastruktur welche aus HPC und Cloud Ressourcen besteht. Auf Basis Web Service und RESTbasierte APIs für Compute Rund Storage Ressourcen angeboten. Diese Infrastrukturservices bieten die Grundlage für generische Datenservices und verteilten Services Wissensmanagement, Wissensentdeckung, Schlussfolgerungen, und semantischen Services. Diese Ebenen bilden die VPH-Share Infostructure und werden für die verteilte und transparente Ausführung von unterschiedlichen Patientenzentrierten Workflows.



Anforderungen

Applikationsspezifische Anforderungen:

- Unterstützung der medizinischen Diagnose, der Behandlung, und der Prävention durch die Integration von diversen Datenquellen und der Ermöglichung von computergestützten Simulationen und Analysen

Soziale Anforderungen:

- Einbindung eines Ethikrats
- Richtlinien für klinische Daten
- Einfache und sichere Verfügbarmachung von verteilten Ressourcen
    - Für (Daten) Lieferanten
    - Für Forscher
    - Für Einrichtungen
    - Für Benutzer

Rechtliche Anforderungen:

- Unterschiedliche rechtliche Basis und Sicherheitsvorkehrungen für unterschiedliche Ressourcen
- Umsetzung auf Basis strenger Privacy und Security Richtlinien
- Komplexe verteilte Security Mechanismen

Big Data Charakteristiken *(Welche Big Data Charakteristiken werden adressiert und in welchem Umfang?)*		
	Volume	Mehr als 680 unterschiedliche Datensets aus verschiedenen Bereichen
	Variety	Komplexität in Bezug auf unterschiedlichste Datentypen und Formate (Bilddaten bis relationale Datenbanken)
	Veracity	Komplexität in Bezug auf integrierte Sicht auf heterogene und verteilt

		gespeicherte Datentypen Semantische Echtzeit-Integration der Daten auf Basis von Ontologien
	Value	Vereinheitlichte und integrierte Sicht (Ontologien) auf verteilte Daten verschiedener Stakeholder sowie die Verwendung von Computer-gestützten Simulationen bringt Mehrwert in der Diagnose, der Behandlung, und der Prävention innerhalb der medizinischen Forschung und für Ärzte Diese Technologien werden für verschiedene innovative Anwendungsfälle nutzbar gemacht: - @neurIST - VPHOP - euHeart - Virolab
Big Data Stack *(Welche Big Data Ebenen und welche Technologien werden adressiert und verwendet?)*	Utilization	Semantische Technologien (Linked Data und Ontologien) für die Wissensaufbereitung und die Integration von unterschiedlichen Datenquellen. Bereitstellung von Visualisierungstools für generiertes Wissen, integrierte Daten, sowie Ergebnisse komplexer datenintensiver Analysen und Simulationen.
	Analytics	Implementierung von medizinischen Simulationen und Analysen
	Platform	Multi-Cloud Plattform auf Basis von OpenSource Toolkits
	Management	Verteilte Cloud Lösung welche eine dezentrale Speicherung der Daten ermöglicht; Bereitstellung von verteilten und skalierbaren Cloud Services für die online Mediation von Daten.
Potenziale	Bereitstellung einer Referenzarchitektur für eine gemeinschaftliche Umgebung welche die Barrieren für die gemeinsame Benutzung von Ressourcen (Date, Rechen, Workflows, Applikationen) beseitigt. - Die gemeinschaftliche und integrierte Bereitstellung von verschiedenen Datenquellen kann die Entwicklung von neuen Forschungsfragen und von innovativen Geschäftsfällen fördern. - Das erstmalige Ermöglichen eines gemeinsamen Zugriffs auf verschiedener klinischer Datenquellen kann die Entdeckung neuer beziehungsweise die Verbesserung bestehender	

	Diagnose-, Behandlungs-, und Präventionsmethoden ermöglichen. - Die innovative Kombination von verteilten Cloud Ressourcen zu einer gemeinschaftlichen europäischen Storage und Compute Cloud kann als Vorzeigebeispiel dienen und Anstöße für weitere Projekte liefern.

4.1.3 Leitprojekt Handel

Es wurde ein Handelsunternehmen untersucht, welches stark auf die Anwendung von Datenbezogenen Diensten setzt. Aus Policy-Gründen kann dieses Unternehmen jedoch nicht genannt werden.

Verbesserung der Wettbewerbsfähigkeit durch die Implementierung von Datenbezogenen Anwendungen	
Domäne	Handel
Projekttyp	Internes Projekt
Projektpartner	
Ansprechpartner	
Projektbeschreibung	Im Handel bestehen vielfältige Ansätze, komplexe Lösungen mit Daten zu verbinden. Hierbei erhofft man sich eine verbesserte Wettbewerbsfähigkeit und besseres Verständnis der Kunden. Im untersuchten Unternehmen wurde folgendes implementiert: • Echtzeitanalyse der Verkäufe einer Filiale. Dem Unternehmen stehen vielfältige Möglichkeiten zur Verfügung jede Filiale in Echtzeit zu beobachten. Hierbei werden verschiedene Dinge überprüft. Wesentlich ist die Beobachtung von Aktionen und die Analyse, wie sich diese pro Filiale unterscheiden. Ist am Wochenende das Produkt „XY“ verbilligt, so kann festgestellt werden wo es sich besser und wo schlechter verkauft und auf „Verkaufsspitzen“ besser reagiert werden. • Auffinden von Anomalien in einer Filiale. Das Unternehmen hat Erfahrungswerte und Muster über Verkäufe einer gewissen Produktkategorie in der Filiale. Ein Muster ist hier zum Beispiel „XY wird alle 15 Minuten verkauft“. Ist dem nicht so, so wird eine Benachrichtigung an FilialmitarbeiterInnen gesendet, mit der Aufforderung das

	jeweilige Produkt zu überprüfen. Es hat sich herausgestellt, das mit dem Produkt jeweils etwas nicht in Ordnung war. • Anpassen der Filialen an demographische Entwicklungen. In bestimmten Fällen ändert sich die Demographie des Einzugsgebietes stark. Dies kann durch die Stadtentwicklung oder aber auch durch Einfluss von Bevölkerungsgruppen der Fall sein. Dem Unternehmen stehen Verfahren zur Verfügung, welche es erlauben, diese zu erkennen und darauf zu reagieren. Eine Reaktion kann z.B. der Umbau der Filialen oder die Auswechselung des Produktsortiments sein.	
Lösungsansatz	Die Anwendungen wurden mit vorhandenen Lösungen für den Handel realisiert. Im Hintergrund stand hierfür eine NoSQL-Datenbank welche vor allem auf dem Echtzeiteinsatz optimiert wurde.	
Anforderungen	Applikationsspezifische Anforderungen: • Near-Realtime Analyse der Filialen bis max. 1 Minute Verzögerung • Speichern großer Datenmengen der Produkte • Mustererkennung der einzelnen Filialen und Produkten • Analyse des Verhaltens der Kunden und Erkennung von Abweichungen • Kombination von verschiedenen Datenquellen über demographische Daten Soziale Anforderungen: • Abspeicherung von Kundenverhalten Rechtliche Anforderungen: • Anonyme Verarbeitung von Kundendaten	
Big Data Charakteristiken *(Welche Big Data Charakteristiken werden adressiert und in welchem Umfang?)*	**Volume**	Es werden viele Daten über Produkte, Muster und dergleichen abgespeichert. Dadurch entsteht ein großes Datenvolumen
	Variety	Daten kommen aus unterschiedlichen Quellen und müssen in ein Endformat für das Unternehmen übertragen werden
	Veracity	Daten über jeweilige Filialen müssen in Echtzeit überwacht werden können. Dadurch ergeben sich Anforderungen an die Geschwindigkeit der Daten
	Value	Die Daten liefern einen wesentlichen Mehrwert für das Unternehmen. Hierbei geht es primär darum, das Produktsegment zu optimieren. Dies steigert Umsatz und Gewinn des

		Unternehmens.
Big Data Stack *(Welche Big Data Ebenen und welche Technologien werden adressiert und verwendet?)*	**Utilization**	Die Analyse Ergebnisse können flexibel in Web-basierte Oberflächen integriert werden. Ziel ist die interaktive Analyse von integrierten Datenquellen.
	Analytics	Die Algorithmen basieren auf statistischen Modellen.
	Platform	Für die Analyse kommen für den Einzelhandel optimierte Systeme zum Einsatz.
	Management	Die Speichersysteme werden durch einen Partner/Outsourcingprovider zur Verfügung gestellt.
Potenziale	Es besteht noch viel Potenzial in Richtung Echtzeitanalyse. Nach aussagen des IT Verantwortlichen des Unternehmens steht dieses erst am Anfang der Möglichkeiten. Hierfür werden in den nächsten Monaten und Jahren viele Projekte stattfinden.	

4.1.4 Leitprojekt Industrie: Katastrophenerkennung mit TRIDEC

Real-Time fähige Anwendung zur Katastrophenerkennung	
Domäne	Industrie
Projekttyp	EU-Projekt
Projektpartner	Joanneum Research
Ansprechpartner	DI Herwig Zeiner Telefon: +43 316 876-1153 Fax: +43 316 876-1191 herwig.zeiner@joanneum.at
Projektbeschreibung	In TRIDEC (09/2010-10/2013) wurden neue real-time fähige Architekturen und Werkzeuge entwickelt, um Krisensituationen zu überwinden und Schäden oder negative Einflüsse nach Möglichkeit durch angepasste Entscheidungen abzuwehren. Zentrale Herausforderung war die Konzeption und Entwicklung einer offenen Plattform für interoperable Dienste, die ein intelligentes, ereignisgesteuertes Management von sehr großen Datenmengen und vielfältigen Informationsflüssen in Krisensituationen ermöglicht. Im Mittelpunkt stand die Arbeit an neuen Modellen und den dazugehörenden real-time fähigen Algorithmen für das Monitoring System. Das Paket beinhaltet auch einen entsprechenden Systemaufbau mit einer sicheren Kommunikations- und

	Analyseinfrastruktur.	
Lösungsansatz	Der technologische Ansatz wurde in zwei Anwendungsfeldern demonstriert, die sich beide durch das Auftreten extrem großer Datenmengen auszeichnen. Das erste Anwendungsszenario setzte den Fokus auf Krisensituationen, wie sie bei der Erschließung des Untergrundes durch Bohrungen auftreten können, einer für Geologen außerordentlich wichtigen, jedoch überaus teuren Aufschlussmethode. Bohrungen werden unter Verwendung von Sensornetzwerken permanent überwacht und Störungen im Bohrbetrieb frühzeitig ermittelt. In TRIDEC wurden vor allem an der Entwicklung neuer Analyseverfahren zur Erkennung und Vermeidung kritischer Situationen bei Bohrungen gearbeitet. Dazu zählt z.B. die Erkennung von Stuck Pipe Situationen genauso wir Vorschläge zur Durchführung krisenvermeidender Operationen (z.B. Ream & Wash).	
Anforderungen	Applikationsspezifische Anforderungen: • Real-Time Analyse von Katastrophenfällen • Real-Time Reaktion auf Katastrophenfälle	
Big Data Charakteristiken *(Welche Big Data Charakteristiken werden adressiert und in welchem Umfang?)*	**Volume**	-
	Variety	-
	Veracity	Industrieanlagen müssen in Echtzeit analysiert werden können. Dies geschieht in Abstimmung mit geografischen Daten.
	Value	Tritt ein Katastrophenfall wie z.B. ein Erdbeben oder ein Zunami ein, so kann frühzeitig darauf reagiert werden und der Schaden an Industrieanlagen minimiert werden.
Big Data Stack *(Welche Big Data Ebenen und welche Technologien werden adressiert und verwendet?)*	**Utilization**	Die Auswertungen werden durch eine graphische Oberfläche repräsentiert.
	Analytics	Die Algorithmen basieren auf statistischen Modellen.
	Platform	-
	Management	-
Potenziale	Hierbei handelt es sich primär um ein Forschungsprojekt. Österreichische Unternehmen, vor allem aus der Öl- und Gasindustrie sowie in produzierenden Unternehmen können dies Anwenden. Ferner besteht die Möglichkeit, ein Produkt daraus zu entwickeln.	

4.1.5 Leitprojekt Weltraum: Prepare4EODC

Projekt Prepare4EODC	
Domäne	ASAP - Austrian Space Applications Programme
Projektpartner	Vienna University of Technology (TU Wien), Department of Geodesy and Geoinformation (GEO) EODC Earth Observation Data Centre for Water Resources Monitoring GmbH (EODC) GeoVille GmbH (GeoVille) Central Institute for Meteorology and Geodynamics (ZAMG) Catalysts GmbH (Catalysts) Angewandte Wissenschaft, Software und Technologie GmbH (AWST)
Ansprechpartner	Mag. Stefan Hasenauer Vienna University of Technology (TU Wien) Tel: +43 1 58801 12241 Email: stefan.hasenauer@geo.tuwien.ac.at DI Dr. Christian Briese EODC Earth Observation Data Centre for Water Resources Monitoring GmbH Tel: +43 1 58801 12211 Email: christian.briese@geo.tuwien.ac.at
Projektbeschreibung	Im Frühjahr 2014 wurde die EODC Earth Observation Data Centre for Water Resources Monitoring GmbH (EODC GmbH) im Rahmen eines „public-private Partnership“ gegründet. Das EODC soll die Nutzung von Erdbeobachtungsdaten für die Beobachtung globaler Wasserressourcen durch eine enge Zusammenarbeit mit Partnern aus der Wissenschaft, der öffentlichen Hand als auch der Privatwirtschaft fördern. Das Ziel dieser Kooperation ist der Aufbau einer kooperativen IT-Infrastruktur, die große Erdbeobachtungsdatenmengen (der neue Sentinel 1 Satellit der ESA liefert freie Daten von ca. 1,8 TB pro Tag) vollständig und automatisch prozessieren kann. Konkret wird das Projekt die folgenden Dienste vorbereiten: 1) Datenzugriff, Weiterverteilung und Archivierung von Sentinel-Daten, 2) eine Sentinel-1 Prozessierungskette, 3) ein Informationssystem für die Überwachung landwirtschaftlicher Produktion basierend auf Sentinel und 4) eine Cloud-Plattform für die wissenschaftliche Analyse von Satellitenbodenfeuchtigkeitsdaten.

Lösungsansatz	Entwicklung von Managementmethoden, grundlegender Software-Infrastruktur sowie ersten Datenverarbeitungsketten.
Anforderungen	Applikationsspezifische Anforderungen: near-realtime processing, advanced processing of large data volumes, long term preservation and reprocessing Soziale Anforderungen: public-private Partnership Rechtliche Anforderungen: Inspire Richtlinien
Big Data Charakteristiken *(Welche Big Data Charakteristiken werden adressiert und in welchem Umfang?)*	1.8 TB pro Tag Disk space: 3-5 PB Anbindung an den Vienna Scientific Cluster (VSC)
Potenziale	Das Ergebnis des Projektes wird positive Auswirkungen auf den Technologievorsprung Österreichs im Bereich der Erdbeobachtung haben. Darüber hinaus stellt die Erfassung globaler Wasserressourcen ein aktuelles Interesse der Gesellschaft dar. Neben nationalen Kooperationen sollen auch internationale Partner in die Aktivitäten des EODC eingebunden werden. Dadurch wird die internationale Rolle Österreichs im Bereich der Nutzung von Satellitendaten gestärkt.

4.2 Implementierung eines Leitfadens für „Big Data“ Projekte

Für die effiziente und problemlose Umsetzung eines Big Data Projekts sind mannigfaltige Aspekte aus vielen unterschiedlichen Bereichen zu beachten welche über normales IT Projektmanagement und diesbezügliche Vorgehensmodelle hinausgehen. Durch die Neuartigkeit der Problemstellungen und der oftmaligen mangelnden unternehmensinternen Erfahrung entstehen zusätzliche Risiken, welche durch ein effizientes Big Data-spezifisches Projektmanagement verringert werden können.

In diesem Leitfaden für die Abwicklung eines Big Data Projekts wird auf die zu beachtenden Spezifika eingegangen und es werden entsprechende Rahmenbedingungen definiert. Der Leitfaden soll Organisationen eine Hilfestellung und Orientierung für die Umsetzung von Big Data Projekten bieten. Der Leitfaden gliedert sich in die Definition eines Big Data Reifegradmodells auf Basis dessen der aktuelle Status bezüglich Big Data in einer Organisation eingeschätzt werden kann und weitere Schritte gesetzt werden können. Des Weiteren wird ein Vorgehensmodell für die Umsetzung von Big Data Projekten vorgestellt welches sich in acht Phasen gliedert. Ein wichtiger Aspekt für die erfolgreiche Umsetzung eines Big Data Projekts ist die Entwicklung von entsprechender Kompetenz innerhalb der Organisation. Hierfür werden benötigte Kompetenzen und Berufsbilder definiert und deren Aufgabenbereiche erläutert. Bei der Speicherung und Verarbeitung von großen Datenmengen

gilt es datenschutzrechtliche Aspekte und die Sicherheit der Daten zu beachten und zu gewährleisten. Hierfür sind grundlegende Informationen in Bezug auf Big Data dargestellt.

4.2.1 Big Data Reifegrad Modell

Der Einsatz von Big Data kann durch schnellere Entscheidungsprozesse als Basis für einen kompetitiven Vorsprung dienen. Dieser steigende Fokus auf Big Data Lösungen bietet einerseits große Gelegenheiten, andererseits ergeben sich dadurch auch große Herausforderungen. Das Ziel ist es nicht nur Informationen abgreifen zu können sondern diese zu analysieren und auf deren Basis zeitnah Entscheidungen treffen zu können.

In vielen Organisationen fehlt es derzeit an der Kompetenz und an dem nötigen Reifegrad um die gesamte Bandbreite von Technologien, Personalbesetzung bis zum Prozessmanagement in Bezug auf Big Data abzudecken. Diese Schritte sind aber für die effiziente und erfolgreiche Ausnutzung der vorhandenen Big Data Assets mit dem Ziel der durchgängigen Umsetzung von Big Data Analysen für die Optimierung von operationalen, taktischen und strategischen Entscheidungen notwendig. Die Fülle an technologischen und analytischen Möglichkeiten, der benötigten technischen und Management Fähigkeiten sowie der derzeit herrschende Hype in Bezug auf Big Data erschweren die Priorisierung von Big Data Projekten innerhalb von Organisationen.

In diesem Big Data Reifegrad Modell werden mehrere voneinander nicht unabhängige Anforderungen berücksichtigt welche die Einordnung einer Organisation in einen bestimmten Reifegrad ermöglichen:

- **Kompetenzentwicklung:** Ein ganzheitlicher Ansatz bezüglich Big Data innerhalb einer Organisation benötigt eine umfassende Strategie zur Entwicklung spezifischer Kompetenzen im Umfeld von Big Data. Diese beinhalten technologische Kompetenzen in Bezug auf den gesamten Big Data Stack, Projektmanagement und Controlling Kompetenzen in Bezug auf Big Data, Kompetenzen in Business Development sowie Know-how bezüglich der Rechtslage und des Datenschutzes. Eine detailliertere Analyse der Kompetenzentwicklung wird in Kapitel 4.2.3 ausgearbeitet.
- **Infrastruktur:** Für die Verarbeitung und Analyse der Daten wird eine Anpassung der IT-Infrastruktur in Bezug auf die vorhandenen Datenquellen, der Anforderungen an die Analyse sowie deren Einbindung in die aktuell verfügbare Systemlandschaft benötigt. Des Weiteren ist für eine effiziente Umsetzung einer Big Data Strategie die Einbindung der unterschiedlichen Stakeholder innerhalb des Unternehmens erforderlich. Hierbei sollen alle Abteilungen welche entweder mit der Datenbeschaffung, Datenaufbereitung und Speicherung sowie alle Abteilungen die Daten für tiefgehende Analysen benötigen eingebunden werden um eine möglichst einheitliche an allen Anforderungen entsprechende Systemlandschaft zu ermöglichen.
- **Daten:** Die Menge an intern sowie extern verfügbaren Daten wächst kontinuierlich und diese sind in den unterschiedlichsten Formaten verfügbar. Neben der Anpassung der Infrastruktur an die gestiegenen Bedürfnisse auf Grund dieser Datenmenge müssen auch weitere Daten-spezifische Aspekte von Datenmanagement, Governance, Sicherheit bis zu Datenschutz berücksichtigt werden.
- **Prozessumsetzung:** Die Umsetzung von Big Data Projekten und der dadurch generierte Erkenntnisgewinn innerhalb einer Organisation wird die Aufdeckung von Ineffizienzen und die Identifizierung von neuen Interaktionsmöglichkeiten mit Kunden, Angestellten, Lieferanten, Partnern und regulatorischer Organisationen ermöglichen. Eine Gefahr

bezüglich der vollständigen Ausschöpfung der Potenziale besteht durch die Möglichkeit einer Organisation die Geschäftsprozesse effizient umzugestalten und an die neuen Erkenntnisse anzupassen. Optimierte Geschäftsprozesse bieten eine höhere Chance in Richtung der schnellen Einführung von innovativen Big Data Anwendungsfällen und Betriebsmodellen.

- **Potenziale:** Aktuell stützen nur wenige Unternehmen ihre Geschäftsprozesse auf eine ganzheitlich definierte Big Data Strategie. Viele Big Data Projekte werden von der IT angestoßen und sind, unabhängig von der technologischen Umsetzung, oft nicht in die bestehende Geschäftsprozesse integriert. Eine enge Vernetzung der Geschäftsziele und Prozesse mit der Umsetzung in der IT kann einen innovativen geschäftsgetriebenen Investmentzyklus in einer Organisation schaffen, welcher die Ausschöpfung der Potenziale ermöglicht. Eine Gefahr besteht in zu hohen Erwartungen in erste Big Data getriebene Projekte durch den aktuell vorhandenen Hype bezüglich dieser Thematik. Diese Risiken können durch die Entwicklung und Umsetzung einer allgegenwärtigen Organisationsweiten Big Data Strategie minimiert werden.

Das hier vorgestellte Reifegradmodell kann als Unterstützung bei der Beurteilung der aktuell vorhandenen Fähigkeiten in Bezug auf die erfolgreiche Umsetzung von Big Data Projekten verwendet werden. Dieses Reifegrad Modell definiert sechs Ebenen welche unterschiedliche Reifegrade in der Umsetzung von Big Data Projekten definieren. Ausgehend von einem Szenario in welchem keinerlei Big Data spezifische Vorkenntnisse in einer Organisation vorhanden sind, erhöht sich der definierte Reifegrad bis zur erfolgreichen Umsetzung eines nachhaltigen Big Data Business innerhalb einer Organisation.

Folgend werden die einzelnen Ebenen im Detail vorgestellt sowie werden Ziele, Wirkungen und Maßnahmen beschrieben die von Organisationen in dieser Ebene gesetzt werden können.

Abbildung 17: Big Data Maturity Modell

4.2.1.1 Keine Big Data Projekte

In diesem Status hat die Organisation noch keine Aktivitäten im Big Data Bereich gestartet. Es wurden noch keine Erfahrungen gesammelt und es werden klassische Technologien für die Datenverwaltung und Analyse eingesetzt.

4.2.1.2 Big Data Kompetenzerwerb

In dieser Phase startet die Organisation erste Initiativen zum Erwerb von Big Data spezifischen Kompetenzen. Die Organisation definiert Early Adopters für den Big Data Bereich. Diese Early Adopters sind für die Sammlung und den Aufbau von Wissen innerhalb der Organisation zuständig. Des Weiteren werden mögliche Einsatzszenarien von Big Data Technologien in Bezug auf die IT Infrastruktur, einzelne Bereiche oder auch bereichsübergreifend entwickelt.

Diese Phase ist oft durch ein sehr hektisches und unkoordiniertes Vorgehen gekennzeichnet. Oft scheitern erste Versuchsprojekte an den verschiedensten Faktoren. Diese sind unzureichende technische Erfahrungen der MitarbeiterInnen, unzureichende Erfahrungen über die Potenziale von Big Data im Unternehmen, fehlendes Bewusstsein über die vorhandenen Daten und wie diese genutzt werden können sowie fehlende Management-Unterstützung. In vielen Fällen werden Projekte nicht zentral gesteuert sondern vielmehr ad-hoc erstellt, da ein gewisser Bedarf besteht. Das zentrale IT-Management ist nicht immer in Kenntnis der Projekte. In dieser Phase besteht eine erhöhte Gefahr, dass Projekte scheitern.

4.2.1.3 Evaluierung von Big Data Technologien

In dieser Phase wurden erste Big Data spezifische Kompetenzen erworben und mögliche Einsatzbereiche dieser Technologien wurden definiert. Die Organisation beginnt den strategischen Aufbau von Big Data Beauftragten. Die Organisation hat mit der Umsetzung von Big Data Projekten zur Optimierung von Geschäftsbereichen oder der IT Infrastruktur begonnen.

Die IT-Verantwortlichen innerhalb des Unternehmens beziehungsweise der Organisation beginnen, das Thema Big Data wesentlich strategischer zu sehen. Es werden erste Strategien und Projektmanagementtechniken evaluiert und umgesetzt. Vielfach handelt es sich jedoch noch um isolierte Versuchsballone, wo zum einem der technologische Kompetenzerwerb erleichtert werden sollte und zum anderen die jeweilige Implementierungsverfahren überprüft werden.

In Österreich findet man sich Großteils in dieser beziehungsweise der ersten Phase wieder. Die Verfasser der Studie konnten mit wichtigen IT-Entscheidungsträgern der heimischen Wirtschaft diesbezüglich sprechen.

4.2.1.4 Prozess für Big Data Projekte

Die initialen Big Data Projekte befinden sich vor dem Abschluss und erste Evaluierungsergebnisse sind vorhanden. Die Big Data Verantwortlichkeiten im Unternehmen sind klar definiert und strategische Maßnahmen für die Umsetzung von weiteren Big Data Projekten und deren Integration in Geschäftsprozesse sind gesetzt. Big Data hat sich in der Geschäftsstrategie der Organisation etabliert.

Hier findet sich auch der erste Ansatz von Top-Management-Unterstützung wieder. Es wurde erfolgreich bewiesen, dass Big Data Projekte einen gewissen Mehrwert im Unternehmen bieten und seitens der Geschäftsführung wird dies anerkannt. Die Big Data Verantwortlichen, welche international auch als „Data Scientists“ bekannt sind, sind nun im Unternehmen vorhanden und arbeiten Strategien und Prozesse für weitere Optimierungen aus. Ein Fokus liegt nun darin, herauszufinden welche weiteren Möglichkeiten es gibt und wie diese erfolgreich in die Tat umgesetzt werden können. Geschäftsprozesse dienen hierbei als wichtiges Mittel, diese Vorhaben zu leiten und zu koordinieren.

4.2.1.5 Gesteuerte Big Data Projekte

Big Data ist fixer Bestandteil der Organisationstrategie und es existiert ein klares Prozessmanagement für Big Data relevante Projekte. Erste Big Data Projekte wurden in Geschäftsprozesse integriert beziehungsweise wurden anhand dieser neue Geschäftsprozesse geschaffen.

Die in der Vorstufe begonnene Standardisierung wird konsequent umgesetzt und erweitert. Die Strategie wird nun ausgebaut und der Schritt zum nachhaltigen Big Data Business wird unternommen.

4.2.1.6 Nachhaltiges Big Data Business

Big Data ist zentraler Bestandteil der Strategie und wird zur Optimierung von unterschiedlichen Geschäftsprozessen eingesetzt. Die IT Infrastruktur setzt Big Data Technologien ein und bietet diese der Organisation für die Umsetzung neuer Projekte an. Big Data spezifische Aspekte sind in das Projektmanagement und die (IT) Strategie des Unternehmens integriert.

Die jeweiligen Big Data Verantwortlichen interagieren mit den unterschiedlichsten Abteilungen des Unternehmens und liefern ständig neue Innovationen. Diese Ebene ist jedoch nicht nur von der Big Data Strategie selbst abhängig sondern benötigt auch eine gewisse Stellung der IT im Unternehmen an sich. Hierfür ist es oftmals notwendig, dass jene Person, welche für die IT im Unternehmen verantwortlich ist, auch Teil der Geschäftsführung ist. Die letzte Stufe kann folglich nur erreicht werden, wenn auch andere Aspekte in diesem Umfang abgedeckt sind.

4.2.2 Vorgehensmodell für Big Data Projekte

Die Definition, Umsetzung und der Einsatz von standardisierten Vorgehensmodellen in IT Projekten ist sehr weit verbreitet. Neben dem Hinweis auf die Wichtigkeit der Umsetzung und Durchsetzung eines geeigneten Vorgehensmodells im Kontext von risikoreichen datengetriebenen Projekten werden diese hier nicht näher beleuchtet. In diesem Abschnitt werden vielmehr essenzielle Aspekte die in direktem Bezug zu Big Data stehen näher betrachtet. Auf deren Basis wird ein Big Data spezifisches Vorgehensmodell vorgestellt, welches während der Organisationsinternen, oder auch Organisationsübergreifenden, Umsetzung von Big Data Projekten unterstützend angewendet werden kann. Die einzelnen Schritte können unabhängig von dem aktuellen Reifegrad in einer Organisation gesehen werden. Das präsentierte Vorgehensmodell muss an die Anforderungen und Größe der Organisation angepasst werden und immer in Bezug auf die vorhandenen und einzubindenden Daten sowie der Unternehmensstrategie gesehen werden.

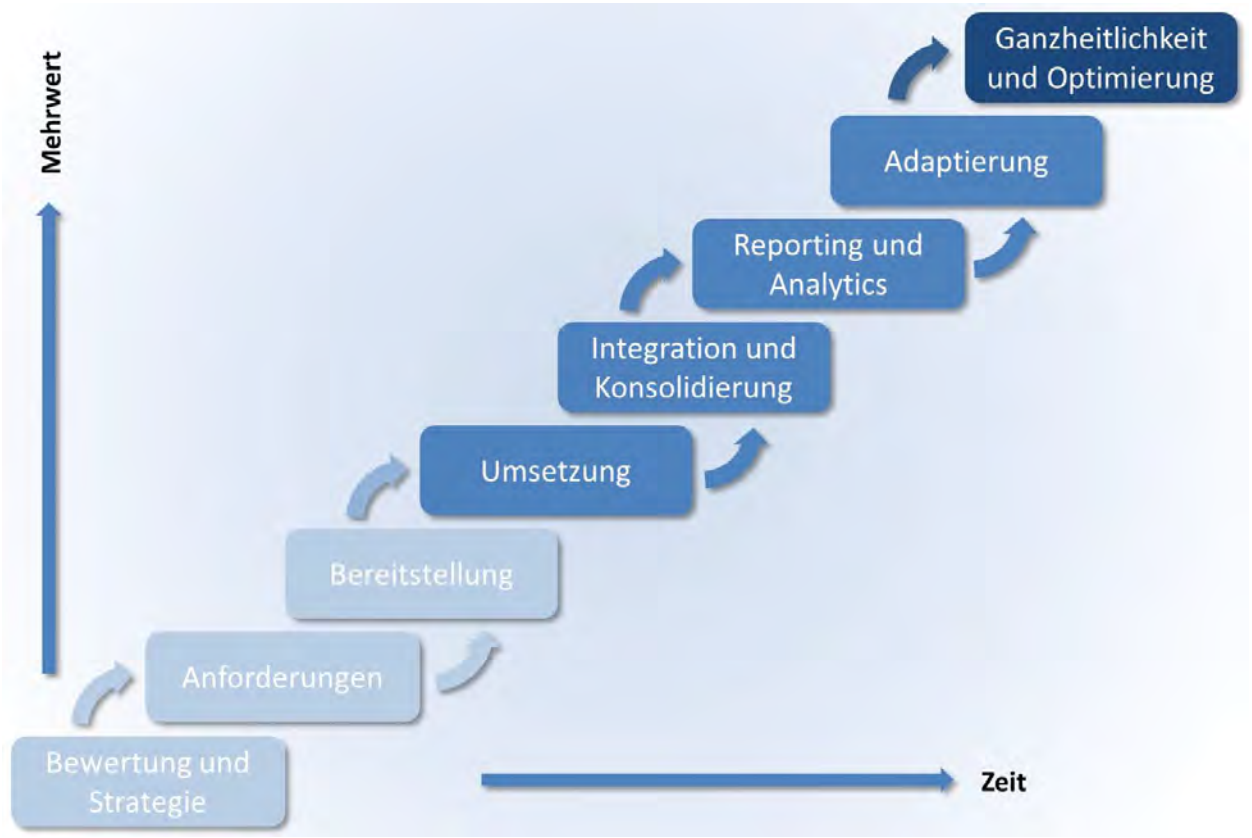


Abbildung 18: Vorgehensmodell für Big Data Projekte

Das Vorgehensmodell besteht aus acht Phasen welche zeitlich nacheinander dargestellt sind, deren Ergebnisse ineinander einfließen, und deren Anwendung von dem gewällten Projektmanagementmodell abhängig ist und dementsprechend angepasst werden müssen. Das Vorgehensmodell (siehe Abbildung 18) gliedert sich in die nachfolgend aufgelisteten Schritte:

- Bewertung und Strategie
- Anforderungen
- Vorbereitung
- Umsetzung
- Integration und Konsolidierung
- Reporting und Analytics
- Adaptierung
- Ganzheitlichkeit und Optimierung

Nachfolgend werden die einzelnen Phasen näher beleuchtet und ihre Spezifika dargestellt.

4.2.2.1 Bewertung und Strategie

Die Phase Bewertung und Strategie ist für die effiziente und erfolgreiche Umsetzung eines Big Data Projektes innerhalb einer Organisation sehr wichtig. Das Ziel dieser Phase ist die Ausarbeitung einer Strategie bezüglich der Umsetzung von Big Data spezifischen Thematiken innerhalb der Organisation. Diese Strategie umfasst Potenziale welche sich aus der Verwendung von Big Data ergeben, klar definierte Ziele in Bezug auf Geschäftsprozesse und Mehrwertgenerierung sowie Vorgehensmodelle, Herausforderungen, benötigte Kompetenzen und Technologiemanagement.

In einem ersten Schritt sollen innerhalb der Organisation die wichtigsten Ziele in Bezug auf Big Data sowie die daraus resultierenden möglichen Potenziale erhoben werden. Ziel hierbei ist es, innerhalb der Organisation Bewusstsein für die Möglichkeiten von Big Data zu schaffen und erste konkrete und machbare Ziele für die potentielle Umsetzung zu generieren.

Auf der Basis dieser Informationen wird in einem nächsten Schritt der aktuelle Status in Bezug auf Big Data erhoben. Dieser Status wird hier wie folgt definiert:

- Identifikation potentieller Geschäftsprozesse
- Identifikation von potentiellem Mehrwert
- Vorhandene interne Datenquellen
- Potentielle externe Datenquellen
- Status der Datenmanagement Infrastruktur
- Status der Datenanalyseinfrastruktur
- Status der Datenaufbereitung und Verwendung
- Vorhandene und benötigte Kompetenzen innerhalb der Organisation
- Analyse der Datenquellen und der Infrastruktur in Bezug auf Privacy und Sicherheit
- Bisherige Projekte im Big Data Bereich
- Big Data Strategie und Big Data in der Organisationsstrategie
- Evaluierung des aktuellen Reifegrades der Organisation in Bezug auf das Big Data Reifegrad Modell

Als Resultat dieser initialen Bewertung soll eine Einschätzung in Bezug auf das Big Data Reifegrad Modell entstehen. Die weiteren Schritte innerhalb der Organisation sollen in Einklang mit den definierten Ebenen des Big Data Reifegrad Modells getroffen werden. Ein weiteres Ergebnis dieser Phase ist eine Bestandsanalyse in Bezug auf Big Data welche bei weiterführenden Projekten nur adaptiert werden muss und eine Grundlage für die effiziente und erfolgreiche Abwicklung eines Big Data Projekts liefert.

Die Ergebnisse der Phase „Bewertung und Strategie“ werden wie folgt zusammengefasst:

- Strategie zu Big Data Themen
- Potenziale und Ziele in Bezug auf Big Data
- Adaptierung der organisationsinternen Strategie in Bezug auf Big Data relevante Themen
- Bewertung der organisationsinternen Umsetzung von Big Data
- Anwendung des Big Data Reifegrad Modells

Für die Durchführung der Phase „Bewertung und Strategie“ können unterschiedliche Methoden angewendet werden. Einige werden hier aufgelistet:

- Interne Workshops: Für die grundlegende Ausarbeitung einer Big Data Strategie und die Erhebung der Potenziale und Ziele wird die Abhaltung eines Big Data Strategie Workshops empfohlen. Ein weiteres Ziel dieses Workshops ist die Spezifikation von organisationsinternen Rollen. Es sollten die internen Big Data Verantwortlichkeiten abgeklärt werden und mindestens eine Person mit der internen Umsetzung und Verwaltung der Big Data Strategie verantwortet werden.
- Interne Umfragen: Anhand von organisationsinternen Umfragen können potentielle Ziele, der aktuelle Stand in Bezug auf Big Data, sowie strategische Optionen erhoben werden.
- Organisationsinterne Interviews: Durch spezifische Interviews zur Big Data Strategie, Potenziale und dem aktuellen Stand der Umsetzung können wichtige Informationen

gesammelt werden. Durch persönlich vom Big Data Beauftragten durchgeführte Interviews kann mit Hilfe der Interviews eine Verankerung des Themas über mehrere Abteilungen hinweg erreicht werden und die Akzeptanz des Themas kann erhöht werden.

- Evaluierung der Marktsituation: Es wird erhoben, welche Big Data Techniken im Markt bereits erfolgreich umgesetzt wurden und was die eigene Organisation daraus lernen kann. Ziel ist es, jene Punkte die erfolgreich umgesetzt wurden zu analysieren und daraus Vorteile abzuleiten. Dies kann auf verschiedenste Art und Weise erfolgen:
    - Auffinden und Analysieren von Use Cases: mit klassischer Desktop-Recherche kann man eine Vielzahl an Big Data Use Cases für verschiedenste Domänen gewinnen. Oftmals sind diese jedoch in anderen geografischen Regionen entstanden, was Einfluss auf die tatsächliche Relevanz des Themas hinsichtlich Datensicherheit und –recht hat.
    - Aufsuchen von Konferenzen und Veranstaltungen: hierbei kann man anhand von erfolgreichen Use Cases lernen. Ferner besteht die Möglichkeit, sich mit Personen anderer Unternehmen auszutauschen.
    - Hinzuziehen von externen BeraternInnen: diese können entweder Branchenfremd oder Branchenaffin sein. Branchenaffine BeraterInnen haben bereits ein dediziertes Wissen über jeweilige Case-Studies, Möglichkeiten und relevanten Technologien. Branchenfremde BeraterInnen haben ein wesentlich breiteres Blickfeld, was unter Umständen neue Ideen bringt und nicht nur das gemacht wird, was andere Marktbegleiter ohnehin schon machen.

Generell wird empfohlen, die Bewertung und Strategie in drei Ebenen durchzuführen:

- Unternehmensinterne Bewertung und Strategie: hier wird vor allem evaluiert, was im Unternehmen vorhanden ist, welche Potenziale sich ergeben und welche Hemmfaktoren die Strategie gefährden könnten.
- Brancheninterne Bewertung und Strategie: In dieser Ebene wird evaluiert, was Marktbegleiter bereits implementiert haben, wo Potenziale gegenüber dem Mitbewerb bestehen und welche Nachteile das Unternehmen gegen eben diesen hat.
- Branchenunabhängige Bewertung und Strategie: In der letzten Ebene wird evaluiert, welche Themen außerhalb der jeweiligen Domäne von Relevanz sind. Dies soll einen wesentlich weiteren Blick auf Big Data Technologien schaffen und die Möglichkeit bringen, Innovation zu fördern.

4.2.2.2 Anforderungen

Die Phase „Anforderungen“ begründet sich auf der erfolgreichen Definition der Big Data Strategie und der umfassenden Bewertung der vorhandenen Big Data Technologien innerhalb einer Organisation. Die Phase Anforderungen ergibt sich aus etablierten IT Vorgehensmodellen mit dem Ziel die projektspezifischen Anforderungen im Detail zu erheben, diese zu analysieren, zu prüfen und mit dem Auftraggeber (intern sowie extern) abzustimmen.

In Bezug auf Big Data Projekte sollte eine erhöhte Aufmerksamkeit auf folgende Punkte gelegt werden:

- Hardware Anforderungen
    - In Bezug auf Big Data Charakteristiken und Big Data Stack
- Software Anforderungen
    - In Bezug auf Big Data Charakteristiken und Big Data Stack

- Funktionale Anforderungen
- Qualitätsanforderungen
- Datenschutz, Privacy, und Sicherheit

Bei der Umsetzung von Big Data Projekten können sich technische Schwierigkeiten in Bezug auf die Big Data Charakteristiken (Volume, Veracity, Velocity) ergeben. Diese stellen sowohl Software als auch Hardware vor neue Herausforderungen und sollten aus diesem Grund gesondert und detailliert behandelt werden. Hierbei müssen aktuell eingesetzte Technologien in Bezug auf deren Einsetzbarkeit mit Big Data evaluiert werden sowie falls notwendig die Anwendbarkeit von neuen Big Data Technologien im Organisationsumfeld bedacht werden.

Des Weiteren ergeben sich durch die Größe, Vielfalt und Geschwindigkeit der Daten additionale Anforderungen und Herausforderungen in Bezug auf die Funktionalität und Qualität. Diese Anforderungen und potentiell daraus resultierende Problematiken sollten ebenfalls vorab gesondert analysiert und evaluiert werden.

Ein wesentlicher Aspekt in Bezug auf die Anwendung von Big Data Technologien in Organisationen ist die Anwendung, Umsetzung und Einhaltung von Datenschutzrichtlinien sowie geeigneten Sicherheitsimplementierungen. Hierzu sollten von Projektbeginn an Datenschutz und Sicherheitsbeauftragte in die Projektabwicklung eingebunden werden. Gerade dem Umgang mit personenbezogenen Daten in Bezug auf Big Data Projekte muss gesonderte und detaillierte Beachtung geschenkt werden. Ebenso wichtig wie die Einhaltung von Datenschutzrichtlinien ist es, sich um gesellschaftliche/soziale Aspekte von Daten zu kümmern. Nur weil etwas nicht explizit durch eine Richtlinie, Verordnung oder geltendes Recht ausgeschlossen ist, heißt es nicht notwendigerweise, das dies unter sozialen Gesichtspunkten auch in Ordnung ist. Hierbei geht es vor allem um die Frage, wie die Organisation mit der Gesellschaft nachhaltig umgeht und wie der Balanceakt zwischen Geschäftsoptimierung mit Daten und der persönlichen Freiheit jedes einzelnen Individuums erfolgreich gemeistert wird.

Für die Analyse von Big Data spezifischen Anforderungen können organisationsinterne Standards und Spezifikationen angewendet werden. Wir schlagen vor folgende Anforderungen in Bezug auf Big Data zu erheben (IEEE Anforderungsspezifikation (IEEE, 1984)):

- Allgemeine Anforderungen
    - Produktperspektive (zu anderen Softwareprodukten)
    - Produktfunktionen (eine Zusammenfassung und Übersicht)
    - Benutzermerkmale (Informationen zu erwarteten Nutzern, z.B. Bildung, Erfahrung, Sachkenntnis)
    - Einschränkungen (für den Entwickler)
    - Annahmen und Abhängigkeiten (Faktoren, die die Entwicklung beeinflussen, aber nicht behindern z.B. Wahl des Betriebssystems)
    - Aufteilung der Anforderungen (nicht Realisierbares und auf spätere Versionen verschobene Eigenschaften)
- Spezifische Anforderungen
    - Funktionale Anforderungen
    - Nicht funktionale Anforderungen
    - Externe Schnittstellen
    - Design Constraints

- Anforderungen an Performance
- Qualitätsanforderungen
- Sonstige Anforderungen

Ein weiterer wichtiger Aspekt in Bezug auf die Erhebung der Anforderungen ist die Erhebung der benötigten Kompetenzen für die erfolgreiche Abwicklung eines Big Data Projekts. Hierbei wird auf Kapitel 4.2.3 verwiesen in welchem die Kompetenzentwicklung in Bezug auf Big Data gesondert behandelt wird.

4.2.2.3 Vorbereitung

Das Ziel der Phase „Vorbereitung“ ist die Anpassung der aktuellen IT Infrastruktur an die Herausforderungen die sich aus der Umsetzung des Big Data Projekts ergeben. Auf Basis der Bewertung der aktuellen IT Infrastruktur in Bezug auf Software und Hardware ist eine umfangreiche Analyse der IT in der Organisation vorhanden. In der Phase „Anforderungen“ wurden detaillierte Anforderungen in Bezug auf die IT Infrastruktur (ebenfalls Hardware und Software) erhoben. In dieser Phase wird der aktuelle Bestand und dessen Leistungsfähigkeiten mit den erhobenen Anforderungen verglichen und eine Gap Analyse (Kreikebaum, Gilbert, & Behnam, 2011) durchgeführt. Diese hat das Ziel die Lücken zwischen IST und SOLL Zustand zu erheben. Des Weiteren bieten diese Ergebnisse die Grundlage für die Planung der weiteren Adaptierung der IT Infrastruktur in Bezug auf Big Data Projekte. Diese weitere Umsetzung sollte immer in Zusammenhang mit der Gesamtstrategie in Bezug auf Big Data gesehen werden und nicht projektspezifisch behandelt werden um mögliche Synergien zwischen Projekten, Anwendungsfeldern sowie Abteilungen optimal ausnutzen zu können.

Hierbei sollen folgenden Punkten zusätzliche Beachtung geschenkt werden:

- GAP Analyse in Bezug auf Big Data Infrastruktur: Hierbei wird analysiert, wie die aktuelle Kapazität aussieht und wie sich diese zukünftig entwickeln wird. Daraus soll sich ergeben, welche Anforderungen für die nächsten Jahre bestehen. Dies hat auch wesentlichen Einfluss auf die IT-Infrastrukturplanung, welche auf externe Speicher (Cloud) oder auch einem Rechenzentrum (intern/extern) ausgelegt werden kann. Neben der Speicherkapazität ist auch die Frage der Analysesysteme relevant. Diese haben andere technische Anforderungen als Speichersysteme. Ein wichtiger Faktor ist die Virtualisierung und Automatisierung der jeweiligen Systeme. Von gehobener Priorität ist die Frage danach, wie flexibel diese Systeme sind.
- GAP Analyse in Bezug auf Big Data Plattformen: Von höchster Priorität ist hier die Analyse der Skalierung und Flexibilität der aktuell und zukünftigen eingesetzten Plattformen. Des Weiteren sollen die jeweiligen Programmiermodelle überprüft werden, ob diese auch dem aktuellen Stand der Technik entsprechen.
- GAP Analyse in Bezug auf Big Data Analytics: Hierbei kommt die Fragestellung zum Tragen, welche Analysesoftware aktuell verwendet wird und wie diese der Strategie für zukünftige Anwendungen dient. Wichtig sind auch die aktuellen Algorithmen und deren Tauglichkeit.
- GAP Analyse in Bezug auf Big Data Utilization: Wichtig ist hierbei, wie sich die jeweiligen Anwendungen mit Big Data Technologien vereinbaren lassen.
- Evaluierung von neuen und vorhandenen Technologien: Damit soll festgestellt werden, wie sich die jeweiligen Technologien für die Umsetzung von Big Data Projekten eignen und wie die Eigenschaften in Bezug auf die Integration in die Systemlandschaft sind.

- Kompetenzen für Umsetzung von Infrastrukturmaßnahmen
- Sicherheit und Datenschutz in der Architektur

4.2.2.4 Umsetzung

Die Phase „Umsetzung“ beschäftigt sich mit der konkreten Implementierung und Integration der Big Data Lösung in die IT Systemlandschaft der Organisation. Bei der Umsetzung des Big Projekts sind neben der Anwendung von etablierten IT Projektmanagementstandards die Beachtung folgender Punkte essenziell:

- Möglichkeit der Integration in bestehende IT Systemlandschaft
- Möglichkeit der Skalierung der anvisierten Lösung in Bezug auf
    - Integration von neuen Datenquellen
    - Wachstum der integrierten Datenquellen
    - Geschwindigkeit der Datenproduktion
- Möglichkeit der Erweiterung der Lösung in Bezug auf
    - Analysealgorithmen für zukünftige Problemstellungen
    - Innovative Visualisierungsmöglichkeiten

Zusätzlich ist die Miteinbeziehung der definierten Big Data Strategie während der konkreten Implementierung des Big Data Projekts essenziell. Ein wichtiger Punkt hierbei ist die Erarbeitung einer organisationsweiten Big Data Infrastruktur welche für konkrete Projekte angepasst beziehungsweise erweitert werden kann. Ein Rahmen für die Umsetzung einer ganzheitlichen Big Data Architektur wird in Kapitel 4.2.5 näher erläutert.

Bereits während der Umsetzung ist ein laufender Soll/Ist Vergleich zu erstellen. Hierbei soll vor allem abgeklärt werden, ob die Ziele mit der umgesetzten Lösung übereinstimmen. Dies dient der Hebung des Projekterfolges. Eine genauere Lösung hierfür wird unter „Ganzheitlichkeit und Optimierung“ dargestellt.

4.2.2.5 Integration und Konsolidierung

Nachdem erfolgreichen Aufbau der (projektspezifischen) IT Infrastruktur und der Umsetzung des Big Data Projekts ist der nächste Schritt die effiziente Integration der (Projekt-) Infrastruktur und Software in die bestehende IT Systemlandschaft. Die vorher geschaffenen Schnittstellen werden genutzt um einen möglichst reibungsfreien Übergang zwischen den Elementen einer umfassenden Big Data Architektur zu schaffen. Des Weiteren werden die umgesetzten und eingesetzten Big Data Plattformen und Tools in die vorhandenen Toolchains eingebunden.

In dem Schritt „Integration und Konsolidierung“ werden nach der Integration der Hardware und Software in die IT Systemlandschaft auch die neuen Datenquellen in das System übergeführt. Ziele hierbei sind einerseits die Bereitstellung der Daten innerhalb der IT Infrastruktur sowie deren Integration in und mit aktuell vorhandenen Datenquellen. Die Art und Weise der Datenintegration hängt sehr stark von der gewählten Big Data Architektur und den Charakteristiken der Daten ab. Für Streaming Daten müssen gesonderte Vorkehrungen getroffen. Ein wichtiger Punkt für die Integration anderer Daten ist die flexible Kombination beziehungsweise die richtungsweisende Entscheidung zwischen klassischen Data Warehousing Zugängen und flexibleren Ansätzen aus dem Big Data Bereich welche die Art und Weise der Datenintegration stark beeinflussen.

Ziel der Integration der gesamten (projektspezifischen) Big Data Technologien in die IT Systemlandschaft ist die Konsolidierung der vorhandenen Systeme und die Schaffung einer ganzheitlichen Big Data Systemlandschaft die einfach für weitere Big Data Projekte genutzt werden kann und auf deren Basis Daten für die Verwendbarkeit in Geschäftsprozessen vorbereitet werden.

Hierbei ist eine elastische Plattform für die jeweiligen Benutzer des Unternehmens von Vorteil. Diese sollte mit wenig Aufwand für diese zur Verfügung stellen und sich am Cloud Computing Paradigma orientieren. Generell soll ein hohes Ausmaß an Self-Services erreicht werden. Hierbei sind mehrere Techniken möglich. Der Fokus soll jedoch auf diesen liegen:

- Infrastructure as a Service: Die Datenplattformen werden anhand von Infrastrukturdiensten zur Verfügung gestellt. Hierbei entfällt für die Benutzer das Management der darunter liegenden Hardware. Einzelne Benutzer müssen jedoch sehr detailliert planen, wie und welche Instanzen diese einsetzen wollen.
- Platform as a Service: Hierbei wird auch der Software-Stack automatisiert. Benutzer bekommen eine flexible Ausführungsplattform zur Verfügung. Damit kann z.B. Hadoop mit nur wenigen Mausklicks bereitgestellt werden und für jeweilige Lasten flexibel skaliert werden. Diese Form bietet die beste Unterstützung der Benutzer, bedeutet jedoch auch mehr Implementierungsaufwand.

4.2.2.6 Reporting und Analytics

Der Schritt „Reporting und Analytics“ befasst sich mit der Implementierung und Bereitstellung von Analysealgorithmen welche auf Basis der vorhandenen Infrastruktur und der vorhandenen Daten umgesetzt werden. Die (projektspezifische) Bereitstellung von Technologien und Methoden (siehe Kapitel 2.2) beeinflusst die Umsetzung der Analyseverfahren maßgeblich. Die Auswahl der verwendeten Technologien beeinflusst hierbei die Möglichkeiten bei der Umsetzung und Entwicklung der Analyseverfahren und sollte aus diesem Grund mit großem Augenmerk, gerade in Bezug auf neue und innovative Szenarien, bedacht werden.

Weiters werden in diesem Schritt spezifische Analysemethoden evaluiert und auf Basis der vorhandenen Technologien umgesetzt und für die Integration in Geschäftsprozesse vorbereitet. Hierbei muss auf die Spezifika der einzelnen betroffenen Geschäftsprozesse sowie auf die Anwendungsdomäne (siehe Domänen in Kapitel 3.2.4) eingegangen werden.

4.2.2.7 Adaptierung

Im Schritt „Adaptierung“ werden die umgesetzten Big Data Technologien auf allen Ebenen des Big Data Stacks in aktuelle und neue Geschäftsprozesse innerhalb des Big Data Projekts angewendet. Durch die Vielfältigkeit der integrierten Daten und die neuartigen Analyseverfahren hat dieser Schritt das Potenzial die etablierten Geschäftsprozesse signifikant und nachhaltig zu verändern. Ziel dieses Schritts ist die Adaptierung der Geschäftsprozesse auf Basis der vorhandenen Datenquellen um diese zu verbessern und Mehrwert zu generieren, also neuen Nutzen (neue Geschäftsfelder, effizientere Geschäftsprozesse, neue Geschäftsmodelle) für die Organisation zu ermöglichen.

4.2.2.8 Ganzheitlichkeit und Optimierung

Nach der erfolgreichen Abwicklung des Big Data Projekts und der entsprechenden Adaptierung der Geschäftsprozesse hin zu effizienteren und neuen Geschäftsmodellen auf Basis von vorhandenen Datenquellen ist es essenziell die gewonnenen Erkenntnisse nachhaltig innerhalb des

Unternehmens zu verankern. Aus diesem Grund werden in dem Schritt „Ganzheitlichkeit und Optimierung“ wichtige Informationen aus der Abwicklung des Big Data Projekts extrahiert und aufbereitet. Ziel hierbei ist es, diese in die vorhandene Big Data Strategie einfließen zu lassen und diese auf Basis der neuen Erkenntnisse zu reflektieren sowie zu erweitern. Diese Erkenntnisse sollen nicht nur in die Big Data Strategie mit einfließen, sondern auch die Bewertung der aktuellen Situation bezüglich Big Data (Big Data Reifegrad Modell und Schritt 1) adaptieren um die Informationen in die Umsetzung von Nachfolgeprojekten einfließen lassen zu können.

Insbesondere werden in diesem Schritt Informationen zu folgenden Punkten gesammelt und in die Big Data Strategie eingearbeitet:

- Vorhandene Datenquellen
    - Typ, Big Data Charakteristiken, Architektur
- Vorhandene Technologien
    - Big Data Stack, Verwendung in Big Data Projekt
- Anforderungen und Herausforderungen
- Potenziale
    - Mehrwertgenerierung
    - Verbesserung von Prozessen
    - Neue Geschäftsprozesse

Durch die Abwicklung und Integration dieses Vorgehensmodells wird die vorhandene Big Data Strategie während der Umsetzung von Projekten verfeinert. Des Weiteren ist ein intaktes Wissensmanagement zu gesammelten Informationen und dem Big Data Reifegrad Modell notwendig und kann die erfolgreiche Umsetzung von Big Data Projekten in Organisationen unterstützen. In diesem Schritt muss auch kritisch reflektiert werden, ob die in den vorherigen Maßnahmen gesetzten Punkte auch zielführend umgesetzt wurden. Hierbei gilt es, Fehler zu analysieren und daraus zu lernen.

4.2.3 Kompetenzentwicklung

Im Zusammenhang mit der wertschöpfenden Umsetzung von Big Data Projekten in Unternehmen und Forschungseinrichtungen entsteht ein großer Bedarf an zusätzlicher Big Data spezifischer Kompetenz um die effiziente und nachhaltige Entwicklung von neuen Geschäftsprozessen voranzutreiben. Die benötigte Kompetenz umfasst den gesamten in 2.1.3 definierten Big Data Stack (Utilization, Analytics, Platform, Management) und es gewinnen neue Technologien, Kenntnisse und Fertigkeiten in diesen Bereichen immer größere Bedeutung. Um Kompetenzen in dem Bereich Big Data innerhalb einer Organisation aufzubauen, können unterschiedliche Maßnahmen gesetzt werden. Gezieltes externes Training von (zukünftigen) Big Data Spezialisten beziehungsweise die Anstellung von neuem in diesem Bereich spezialisiertem Personal können hierbei den ersten wichtigen Schritt setzen. Durch ein vielfältiges und spezialisiertes Angebot an internen Schulungsmaßnahmen kann das Wissen innerhalb einer Organisation verbreitet und vertieft werden.

Grundlage für eine wertschöpfende Umsetzung von Big Data ist die grundsätzliche Verfügbarkeit von hoch qualifiziertem Personal welche innerhalb der Organisation als „early adopters“ eingesetzt werden können. Hierfür bedarf es gerade für einen hochtechnologischen Bereich wie Big Data an grundlegend und wissenschaftlich fundiert ausgebildeten Fachkräften in allen Bereichen des Big Data Stacks. An dieser Stelle der Studie wird auf die Analyse der derzeit angebotenen tertiären Ausbildungen in Österreich (siehe Kapitel 2.4.2) verwiesen.

In den geführten Interviews und den abgehaltenen Workshops wurde von Seiten der Teilnehmer auf den grundsätzlichen Bedarf und ein derzeit zu geringes Angebot an hochqualifiziertem Personal in diesem Bereich hingewiesen. Auf Grund dieser Erfahrungen und dem international ersichtlichen Trend zu dem neuen Berufsbild „Data Scientist“ wird dieser Bereich nachfolgend näher beleuchtet.

4.2.3.1 Data Scientist

Das Berufsbild des Data Scientist hat in der letzten Zeit eine immer größere Bedeutung erlangt. In (Davenport & Patil, 2013) wird dieses Berufsbild als das attraktivste des 21ten Jahrhunderts beworben und in vielen Medien wird auf die Wichtigkeit gut ausgebildeter Fachkräfte für den Bereich Big Data hingewiesen (Bendiek, 2014), (Fraunhofer, 2014). Hierbei wird auch auf das Zitat von Neelie Kroes *„We will soon have a huge skills shortage for data-related jobs.“* (Speech – Big Data for Europe, European Commission – SPEECH/13/893 07/11/2013) hingewiesen. Ein weltweiter Anstieg an Ausschreibungen für dieses neue Berufsbild ist im letzten Jahr zu vermerken und das Berufsbild wird auf unterschiedliche Arten beschrieben. Die Beschreibungen des Berufsbilds Data Scientists reichen von Personen welche Wissen und Methodik aus Analytik, IT und dem jeweiligen Fachbereich vereinigen, über Personen welche Daten in Unternehmen durch die Analyse mit wissenschaftlichen Verfahren und der Entwicklung von prädiktiver Modellen schneller nutzbar machen.

Beispiele der detaillierten Diskussionen der benötigten Fachkenntnisse beschreiben Data Scientists als eine Rolle die aus der Weiterentwicklung von Business oder Data Analysts hervorgeht. Hierfür werden detaillierte Kenntnisse in den Bereichen IT, der Applikation, der Modellierung, der Statistik, Analyse und der Mathematik benötigt (IBM, 2014). Ein weiterer wichtiger Punkt ist die Fähigkeit Erkenntnisse sowohl an leitende Stellen in Business und IT weitervermitteln zu können. Dementsprechend werden Data Scientists als eine Mischung aus Analysten und Künstlern beschrieben. In (KDNUggets, 2014) werden die benötigten Fähigkeiten von Data Scientists unter anderem folgendermaßen beschrieben: Es werden Kenntnisse in den Bereichen statistischer Analyse, High Performance Computing, Data Mining und Visualisierung benötigt. Im Speziellen werden Informatikkenntnisse über Mathematik, Mustererkennung, Data Mining, Visualisierung, Kenntnisse über Datenbanken, im speziellen Data engineering und data warehousing sowie entsprechendes Domänenwissen und unternehmerischer Scharfsinn gefordert.

In (Mason, 2014) wird der Bereich Data Scientist ebenfalls analysiert. Hier wird das von Organisationen geforderte Profil als Kombination aus Informatik, Hacking, Engineering, Mathematik und Statistik dargestellt und kritisch hinterfragt ob diese Kombination möglich ist *(„data scientists are awesome nerds“*). Eine weitere Diskussion des Berufsbilds wird auf (EMC2, 2014) dargestellt. Hier wird auf die innovative Kombination von quantitativen, technischen und kommunikativen Fähigkeiten mit Skepsis, Neugier und Kreativität verwiesen und somit der wichtige Aspekt der sozialen Fähigkeiten hervorgehoben.

Auf Basis breiter Diskussionen über das Berufsbild Data Scientist und Gesprächen im Rahmen der Studienworkshops, Interviews mit Stakeholdern, sowie vorhandenen Datengrundlagen auf Grund von IDC internen Studien und verfügbaren BITKOM Studien (BITKOM, 2013) wird das Berufsbild hier näher klassifiziert. Auf Grund der breiten Masse an geforderten Fähigkeiten ist die Vereinigung dieser in einer Person schwer zu erreichen.

In Tabelle 9 wird ein Überblick über wichtige Kompetenzen für das Berufsbild Data Scientist gegeben:

Technische Kompetenzen	Soziale Kompetenzen	Wirtschaftliche Kompetenzen	Rechtliche Kompetenzen
Utilization - Visualisierung - Grafikdesign - Wissensmanagement	**Teamarbeit**	**Identifikation von innovativen Geschäftsmodellen**	**Datenschutz**
Analysis - Innovative Verknüpfung von Daten - Machine Learning - Statistik und Mathematik	**Kommunikation**	**Evaluierung und Umsetzung von innovativen Geschäftsmodellen**	**Ethik**
Platform - Skalierbare Programmierung - Datenmanagement	**Konfliktmanagement**	**Projektmanagement**	
Management - Infrastrukturentwicklung - Infrastrukturbereitstellung - Skalierbarer Speicher - Massive Infrastrukturen	**Führung**	**Projektcontrolling**	

Tabelle 9: Kompetenzen Big Data Scientist

Auf Basis der erhobenen Kenntnisse wird das Berufsbild näher kategorisiert. Hierbei werden fünf spezifische Kategorien analysiert und in Abbildung 19 deren benötigte Fähigkeiten visualisiert. Unabhängig von der jeweiligen Kategorie wird auf die Wichtigkeit der Einbindung von domänenspezifischem Wissen hingewiesen. In dieser Studie unterscheiden wir zwischen folgenden Berufsbildern:

- Big Data Business Developer
- Big Data Technologist
- Big Data Analyst
- Big Data Developer
- Big Data Artist

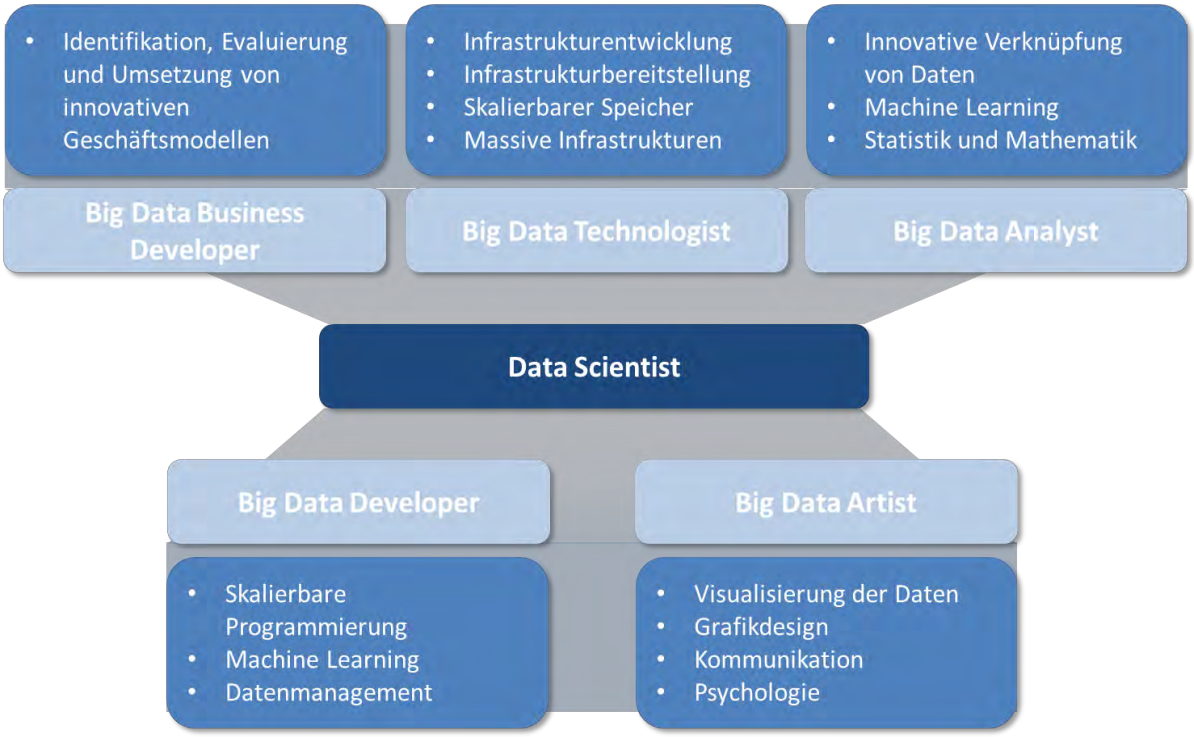


Abbildung 19: Das Berufsbild des Data Scientist

Big Data Business Developer

Das Berufsbild Big Data Business Developer setzt sich aus Kompetenzen in der Mehrwertgenerierung auf Basis von Daten und der Entwicklung aktueller und zukünftiger Geschäftsmodelle zusammen. Einerseits ist das Ziel eines Big Data Business Developers mithilfe von interner Kommunikation neue und innovative Geschäftsfelder und Möglichkeiten auf Basis der vorhandenen und zu integrierenden Daten aufzuspüren und diese gemeinsam mit Mitarbeitern zu entwickeln. Andererseits ist es dessen Aufgabe die Kundenbedürfnisse abzuschätzen, Kontakte zu Daten Providern, Partnern und Kunden zu knüpfen um maßgeschneiderte und innovative Daten-getriebene Produkte zu entwickeln. Demnach ist das Berufsbild als Schnittstelle zwischen der Ebene Utilization und dem Kunden zu sehen.

Big Data Technologist

Das Berufsbild des Big Data Technikers beschäftigt sich mit der Bereitstellung einer für Big Data Szenarien nutzbaren Infrastruktur für die Entwicklung neuer und innovativer Geschäftsmodelle. Hierfür sind breite Kenntnisse von Management und auch Platform Technologien erforderlich. Diese Kenntnisse umfassen die Entwicklung und Instandhaltung von großen Datenzentren, die Verwaltung von Big Data spezifischen Softwarelösungen für die skalierbare Speicherung von großen Datenmengen und auch die Bereitstellung von skalierbare Ausführungsumgebungen für Big Data Analysen.

Big Data Developer

Das Berufsbild des Big Data Entwicklers umfasst die skalierbare Implementierung von Big Data Analysen auf Basis von massiv parallelen Infrastrukturen und Plattformen. Hierfür werden herausragende Kenntnisse in der Parallelisierung von Programmen aus dem Bereich High Performance Computing sowie der skalierbaren Speicherung, effizienten Abfrage von Daten und der Abfrageoptimierung aus dem Datenbankbereich benötigt. Dieser neue Entwicklertypus erfordert eine

grundlegende Informatikausbildung in diesen Bereichen und eine hohe Flexibilität gegenüber neuen Technologien.

Big Data Analyst

Das Berufsbild des Big Data Analysten beschäftigt sich mit dem Auffinden von neuen Verknüpfungen und Mustern in Daten. Hierfür werden fundierte Kenntnisse in den Bereichen Machine Learning, Mathematik und Statistik benötigt. Big Data Analysten entwickeln mathematische oder statistische Modelle, setzen diese mit Hilfe von Big Data Entwicklern um, und wenden diese auf große Datenmengen an um neue Zusammenhänge und Informationen zu generieren.

Big Data Artist

Das Berufsbild des Big Data Artist ist für die visuelle Darstellung und Kommunikation des Mehrwerts für den Endbenutzer und den Kunden zuständig. Hierfür werden einerseits fundierte Kenntnisse in den Bereichen skalierbare Visualisierung, Grafikdesign und Human Computer Interaction für die computergestützte Darstellung der Informationen benötigt. Des Weiteren sollten Big Data Artists eine gute Ausbildung in den Bereichen Kommunikation und Psychologie besitzen um den Effekt der Darstellungsform auf das Gegenüber abschätzen zu können und in das Design einfließen lassen zu können.

4.2.4 Datenschutz und Sicherheit

Datenschutz und Sicherheit sind in Bezug auf Big Data ein sehr wichtiges Thema. In der Umsetzung einer Big Data Strategie, insbesondere in der Umsetzung von Big Data Projekten, muss diesen Themen eine große Bedeutung beigemessen werden. Datenschutz und Sicherheit umfassen mehrere Dimensionen und diese müssen vor und während der Umsetzung sowie während des Betriebs betrachtet werden. Diese Dimensionen reichen von rechtlichen Fragestellungen aus datenschutzrechtlicher Sicht, Sicherheitsstandards welche für die Umsetzung eines Projektes eingehalten werden müssen, bis zu technischen Umsetzungen des Datenschutzes und der Sicherheit der Infrastruktur. Gerade in Bezug auf Big Data Projekte erlangt die Durchsetzung und Beachtung von Datenschutz enorme Bedeutung. Aktuelle Technologien ermöglichen die Speicherung, Verknüpfung, und Verarbeitung von Daten in noch nie dagewesenen Ausmaßen. In diesem Bereich muss auf die Balance zwischen technologischen Möglichkeiten, rechtlichen Rahmenbedingungen, und persönlichen Rechten auf Datenschutz hingewiesen werden und diese muss in weiterer Folge aufrechterhalten werden. Technologische Umsetzung von Datenschutz und Sicherheit sind grundsätzlich vorhanden doch es wird auf die große Bedeutung der Weiterentwicklung von diesen technologischen Möglichkeiten in Bezug auf die weitere Vernetzung von Daten sowie der rechtlichen Rahmenbedingungen hingewiesen.

In Bezug auf Datenschutz und Sicherheit wurden mehrere Studien in Österreich durchgeführt. In mehreren Leitfäden wird die Situation in Bezug auf Datenschutz und Sicherheit in Österreich von diesbezüglichen Experten im Detail ausgearbeitet und der Öffentlichkeit zur Verfügung gestellt. In diesem Kapitel wird aus diesem Grund auf vorhandene Analysen dieses Bereichs verwiesen und auf die Wichtigkeit von Datenschutz und Sicherheit und diesbezüglichen definierten rechtlichen Rahmenbedingungen im Bereich Big Data verwiesen. Für eine detailliertere Analyse auf wird auf folgende frei zugängliche Leitfäden in Bezug auf Österreich der EuroCloud[44] und des IT-Cluster Wien[45] sowie auf den BITKOM Leitfaden für Deutschland verwiesen:

[44] http://www.eurocloud.at/

- EuroCloud, Leitfaden: Cloud Computing: Recht, Datenschutz & Compliance, 2011 (EuroCloud, 2011)
- IT Cluster Wien, Software as a Service – Verträge richtig abschließen 2., erweiterte Auflage, 2012 (IT Cluster Wien, 2012)
- BITKOM, Leitfaden: Management von Big-Data-Projekten, 2013 (BITKOM, 2013)

4.2.5 Referenzarchitektur

Für die effiziente Umsetzung eines Big Data Projekts innerhalb einer Organisation ist die Wahl einer geeigneten Architektur essenziell. Neben den zahlreichen verfügbaren Technologien und Methoden sowie der großen Anzahl an verfügbaren Marktteilnehmer (siehe Kapitel 0) sind die Spezifika der verfolgten Business Cases und deren Anforderungen sowie Potenziale in der jeweiligen Domäne für die Implementierung einer geeigneten Architektur, der entsprechenden Frameworks sowie deren Anwendung in der Organisation notwendig. Um diese wichtigen Entscheidungen innerhalb einer Organisation zu unterstützen werden hier zuerst unterschiedliche Fragestellungen für die Wahl der Architektur diskutiert und im Anschluss wird eine Referenzarchitektur für Big Data Systeme für alle genannten Szenarien definiert.

Um die richtigen technologischen Lösungen für Big Data Projekte zu wählen sollten davor einige Fragestellungen geklärt werden. Derzeit ist hierbei die größte Herausforderung die Wahl der technologischen Plattform unter Rücksichtnahme der Verfügbarkeit von Kompetenzen innerhalb der eigenen Organisation. Wie in (Big Data Public Private Forum, 2013) dargestellt gibt es derzeit eine Vielzahl an Diskussionen bezüglich Anforderungen an Technologien welche zu einem großen Teil aus den verfügbaren Daten resultieren. Darüber hinaus werden aber die benötigten Kompetenzen (siehe Kapitel 4.2.1.2) und das Verständnis für die Herausforderungen und Technologien zu wenig beleuchtet und sind nicht ausreichend vorhanden. Aus diesem Grund hat Yen Wenming in (Yen, 2014) folgende drei Schlüsselfragen definiert und eine grundlegende Kategorisierung von Problemstellungen und Architekturen dargestellt.

- Daten sollen für die Lenkung von Entscheidungen verwendet werden und nicht für deren grundsätzliche Verfügbarkeit gespeichert und analysiert werden.
- Die Analysemethoden sollen regelmäßig an die aktuellen Bedürfnisse angepasst werden.
- Automatisierung der Prozesse ist eine essenzielle Herausforderung um mehr Experimente und Analysen ausführen zu können. Dies ermöglicht die effiziente Betrachtung neuer Fragestellungen und hebt dadurch das Innovationspotenzial.

Des Weiteren wird in Abbildung 20 eine Referenzarchitektur für Big Data Systeme auf Basis des Big Data Stacks dargestellt. Die vorgestellte Architektur umfasst den gesamten Daten Lebenszyklus innerhalb einer Organisation und bezieht auch bestehende Systeme (z.B. ERP Systeme, Relationale Datenbanken, Data Warehouse) mit ein.

[45] http://www.clusterwien.at/overview/de/

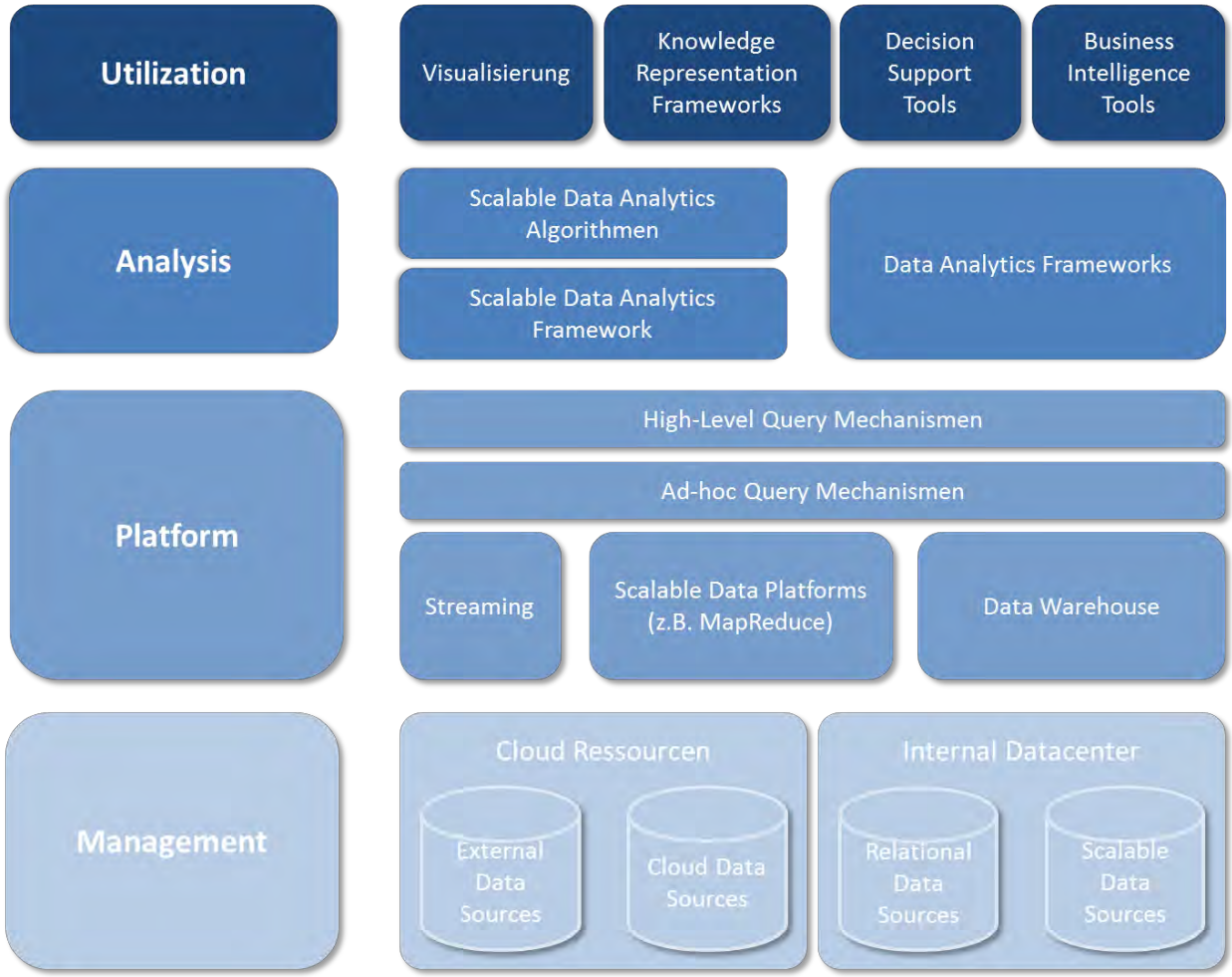


Abbildung 20: Referenzarchitektur für Big Data Systeme

Management Ebene

Die Management Ebene umfasst die Bereitstellung der Infrastruktur sowie die Datenquellen an sich. Die bereitgestellte Infrastruktur wird hierbei in Cloud-basierte Ressourcen und interne Datenzentren unterteilt. Unter Cloud-basierten Systemen werden hierbei einerseits klassische Cloud Ressourcen wie elastische Rechen- und Speicherressourcen verstanden welche von einer Organisation explizit für die Speicherung oder die Berechnung angemietet werden. Andererseits werden in diese Kategorie auch externe Datenquellen welche nach dem Everything as a Service Prinzip (Banerjee & et. al., 2011) angeboten werden eingeordnet. Dies beinhaltet, unter anderem, externe Sensor Systeme (z.B. GPS, Mobilfunk, Wettersensoren) sowie soziale Medien (z.B. soziale Netzwerke und Nachrichtendienste) welche als Service im Internet verfügbar sind. Unter internen Datenzentren werden alle unternehmensinternen Serverlandschaften unabhängig von deren Größe zusammengefasst. Hierbei kann es sich um explizite Big Data Datenzentren großer Firmen wie auch um einzelne Datenbankserver handeln welche in den Big Data Lebenszyklus eingebunden werden. Auf Basis dieser Rechen- und Speicherressourcen werden Daten in unterschiedlichen Formaten und Größen abgespeichert und hierfür werden diverse Systeme verwendet. Für die effiziente Umsetzung eines Big Data Systems werden die Installation oder Anmietung/Einmietung eines Datenzentrums auf Basis von Commodity Hardware (dessen Größe abhängig von der Organisation und der Datenmenge ist) und die Verwendung eines einfachen skalierbaren Speichersystems empfohlen. Hierfür werden von unterschiedlichen Anbietern auch spezielle vorgefertigte Lösungen angeboten

(siehe Kapitel 2.3). Gemeinhin wird eine Erweiterung der bestehenden IT Landschaft mit Big Data Infrastruktur und die flexible Integration der bestehenden Datenquellen empfohlen.

Ein weiterer essenzieller Punkt auf der Management Ebene ist die effiziente und verteilte Speicherung von großen Datenmengen auf Basis der zur Verfügung stehenden Ressourcen. Eine große Bedeutung haben hierbei verteilte Dateisystemen wie zum Beispiel das Hadoop Distributed File System (HDFS) erlangt. Diese ermöglichen die transparente Daten-lokale Ausführung von Plattform Lösungen. Neben verteilten Dateisystemen können aufkommende NoSQL Systeme für die verteilte und skalierbare Speicherung von großen Datenmengen verwendet werden. Diese bieten eine höhere Abstraktion der Daten bei hoher Skalierbarkeit. Neben diesen Big Data Systemen wird das Daten Ökosystem durch relationale Datenbanken ergänzt.

Plattform Ebene

Die Plattform-Ebene beschäftigt sich mit der effizienten Ausführung von Datenanalyseverfahren auf großen Datenmengen wofür massiv parallele, skalierende und effiziente Plattformen benötigt werden. Die Wahl der richtigen Architektur innerhalb eines Projekts hängt stark von den konkreten Fragestellungen und der Charakteristiken der typischen Datenanalysen ab. Diese werden in (Yen, 2014) im Zuge von Big Data häufig in „Batch Systeme“, „Interaktive Systeme“ und „Stream Verarbeitung“ unterteilt. Tabelle 10 gibt einen Überblick über die Charakteristiken dieser Architekturen auf Basis dieser Einteilung. Die Systeme werden anhand der Dauer von Abfragen, des unterstützten Datenvolumens, des Programmiermodells und der Benutzer verglichen. Während Systeme für die Batch Verarbeitung für längere Analysen auf sehr großen Datenmengen welche Ergebnisse nicht sofort zur Verfügung stellen müssen eingesetzt werden können, werden interaktive Systeme meistens auf geringeren Datenmengen, dafür aber mit kurzen Zugriffszeiten, eingesetzt. Systeme für die Echtzeit-Verarbeitung von Datenstreams werden gesondert angeführt. Durch die neuesten Entwicklungen im Bereich der Big Data Systeme (z.B. Apache YARN und Apache Spark) verschwimmen die Grenzen zwischen Batch Verarbeitung und Interaktiven Systemen immer mehr. Ziel für die Zukunft ist es hier gemeinsame Systeme für unterschiedliche Anwendungsfälle auf Basis derselben Daten zu entwickeln. Diese Systeme unterstützen die Ausführung von unterschiedlichen Programmiermodellen welche für differenzierte Szenarien gedacht sind.

	Batch Verarbeitung	Interaktive Systeme	Stream Verarbeitung
Abfragedauer	Minuten bis Stunden	Millisekunden bis Minuten	Laufend
Datenvolumen	TBs bis PBs	GBs bis PBs	Durchgängiger Stream
Programmiermodell	MapReduce, BSP	Abfragen	Direkte Azyklische Graphen
Benutzer	Entwickler	Entwickler und Analysten	Entwickler

Tabelle 10: Big Data Architekturen (Kidman, 2014)

In der beschriebenen Referenzarchitektur sind diese Systeme auf Grund dieser Entwicklungen anders eingeteilt. Streaming Lösungen werden für die Echtzeitanbindung von (externen) Datenquellen sowie Sensorsystemen benötigt. Diese Lösungen können Daten in Echtzeit analysieren beziehungsweise die vorverarbeiteten Daten in das Big Data System einspeisen. Skalierbare Datenplattformen verfolgen massiv parallele Programmierparadigmen und werden auf Basis von internen Datenzentren (oder von Cloud Ressourcen) bereitgestellt. Drittens umfasst dieser Bereich auch klassische Data Warehouses welche in Organisationen verfügbar sind. Im Unterschied zu der vorhergehenden Einteilung (siehe Tabelle 10) sind Ad-Hoc und High-Level Abfragesprachen über den skalierbaren Plattformen und Date Warehouses angeordnet. Diese können als zusätzliche Abstraktionsschicht für die einfachere Benutzung und den interaktiven Zugriff auf dieselben Daten verwendet werden.

Analytics Ebene

Der Bereich Analytics beschäftigt sich mit der Informationsgewinnung aus großen Datenmengen auf Basis von mathematischen Modellen, spezifischen Algorithmen oder auch kognitiven Ansätzen. Das Ziel hierbei ist es unter anderem, neue Modelle aus Datenmengen zu erkennen, vorhandene Muster wiederzufinden oder neue Muster zu entdecken. Auf dieser Ebene werden Methoden aus dem Machine Learning, der Mathematik oder der Statistik angewendet. Hierfür stehen Organisationen unterschiedlichste Technologien von kommerziellen Betreibern als auch aus dem OpenSource Bereich bereit. Die größte sich hier stellende Herausforderung ist die Anpassung der Algorithmen an massiv parallele Programmiermodelle der Plattform Ebene um diese auf riesigen Datenmengen skalierbar ausführen zu können.

In der Referenzarchitektur werden drei Subbereiche dargestellt: Data Analytics Frameworks, Scalable Data Analytics Frameworks und Scalable Data Analytics Algorithms. In Organisationen werden häufig klassische Data Analytics Frameworks produktiv in den unterschiedlichsten Bereichen eingesetzt. Derzeitiges Ziel der Hersteller ist es diese Frameworks an die Herausforderungen im Bereich Big Data anzupassen. Dies geschieht meistens durch die nahtlose Unterstützung und Integration von unterschiedlichen massiv parallelen Programmiermodellen. Da es sich hierbei um einen wichtigen aber derzeit noch nicht vollständig abgeschlossenen Prozess handelt werden diese Frameworks gesondert dargestellt. Neben klassischen Data Analytics Frameworks entstehen immer mehr neue Frameworks welche speziell für die massiv parallele Ausführung von Datenanalysen entwickelt werden. Hierbei entstehen abstrahierte

Programmiermodelle welche die Parallelisierung (teilweise) vor den Anwendern verstecken können und gleichzeitig hochperformante Bibliotheken für Machine Learning, Mathematik und Statistik in das Framework einbauen. Die Entwicklung dieser Frameworks ist derzeit ein aufstrebender Bereich und gerade diese werden den Bereich Big Data in den nächsten Jahren vorantreiben.

Algorithmen für spezifische Problemstellungen werden auf Basis der zur Verfügung stehenden Frameworks für Datenanalyse entwickelt und bereitgestellt. Hierbei ist auf die Anwendbarkeit der Algorithmen auf große Datenmengen und der möglichst transparenten massiven Parallelisierung dieser zu achten. Mittlerweile existieren einige skalierbare Machine Learning Frameworks welche ausgewählte Algorithmen in skalierbarer Form bereitstellen. Generell ist für die Analyseebene ein breites Spektrum an Kompetenz (richtige Algorithmen, massiv parallele und skalierbare Implementierungen, Verwendung von Big Data Frameworks) von Nöten und dies ist auch in der Auswahl der Tools zu beachten.

Utilization Ebene

Der Einsatz von Big Data-Technologien und -Verfahren wird von Unternehmen und Forschungseinrichtungen unabhängig von dem Geschäftsfeld mit einem bestimmten Ziel angedacht: Big Data-Technologien sollen Mehrwert generieren und dadurch die aktuelle Marktsituation der Organisation stärken. Utilization Technologien bilden diese Schnittstelle zwischen dem Benutzer und den darunter liegenden Technologien. Hierbei wird im Rahmen der Referenzarchitektur einerseits auf die notwendige Integration in die vorhandene Toolchain in Bezug auf Wissensmanagement, Business Intelligence und entscheidungsunterstützende Systeme verwiesen. Des Weiteren kann die Integration von Wissensrepräsentationssystemen, skalierbaren und interaktiven Visualisierungssystemen sowie von Visual Analytics als einfache und interaktive Schnittstelle für Benutzer das Potenzial von Big Data Technologien innerhalb der Organisation enorm heben.

5 Schlussfolgerungen und Empfehlungen

Big Data ist ein international florierender Bereich der enorme wirtschaftliche und technologische Potenziale in sich birgt. In der Wissenschaft kann dahingehend ein Trend zu dem vierten Paradigma (Hey, Tansley, & Tolle, The Fourth Paradigm: Data-Intensive Scientific Discovery, 2009) der datenintensiven Forschung beobachtet werden welches neben empirischen Untersuchungen und Experimenten, analytischen und theoretischen Herangehensweisen und rechenintensiver Wissenschaft sowie Simulationen entsteht. Dies beinhaltet die neuartige Nutzung von vorhandenen und entstehenden riesigen Datenarchiven auf deren Basis mit wissenschaftlichen Methoden neue Forschungsfragen untersucht werden können. Hierfür werden neue Technologien und Methoden im Bereich datenintensiver Programmiermodelle, Infrastrukturen, analytischen Methoden bis zu Wissensmanagement notwendig, welche sich in den letzten Jahren enorm weiterentwickeln. Die Weiterentwicklung von diesen Methoden wird stark von Anwendungsseite getrieben und in den unterschiedlichsten Bereichen wie zum Beispiel Medizin, Raumfahrt, Verkehr und Physik erfolgreich eingesetzt um neue Erkenntnisse zu generieren und diese Bereiche voranzutreiben. Weiters werden für deren effiziente Umsetzung große Anstrengungen gerade im Bereich der skalierbaren Speichersysteme und Programmierparadigmen gesetzt. Diese bieten enormes Potenzial und sind für den Erfolg und für die Entstehung innovativer Produkte im Big Data Bereich essenziell. Andererseits finden diese Technologien in der Wirtschaft immer mehr praktische Umsetzungen und werden von vielen Unternehmen erfolgreich eingesetzt, um vorhandene Potenziale zu heben, sich einen Marktvorteil zu verschaffen und um neue Geschäftsfelder zu erschließen.

Die vier wesentlichen Voraussetzungen für eine erfolgreiche Umsetzung von Big Data Projekten sind einerseits die Verfügbarkeit der Daten und deren Güte, die Rechtsicherheit bei der Verwendung der Daten, das zur Verfügung stehen der benötigten Infrastruktur sowie das Vorhandensein der benötigten Kompetenz auf allen vier Ebenen des Big Data Stacks (Utilization, Analytics, Platform, Management), welche für die Erschließung neuer Erkenntnisse auf Basis der Daten erforderlich ist. Diese Voraussetzungen können in dem folgenden Motto zusammengefasst werden:

„Daten sind der Rohstoff – Kompetenz ist der Schlüssel“

Eine wesentliche Erkenntnis aus den durchgeführten Interviews und Workshops ist, dass es in Bezug auf Daten oftmals klarere Rahmenbedingungen für deren Nutzung bedarf, vor allem wenn es sich um sensible Daten (bspw. mit indirektem Personenbezug) handelt. Die aktuelle Situation wirkt derzeit in vielen Unternehmen und Forschungsprojekten als Hemmschuh für den effizienten Einsatz von innovativen Szenarien (siehe Kapitel 5.2.1 sowie 5.2.2). Als positives Beispiel wird hier auf Open Data Initiativen (z.B. Bereitstellung von Verkehrsdaten), sowie auf die Etablierung von Datenmarktplätzen in Österreich/Europa verwiesen. Hierbei werden Daten und deren Nutzung unter klaren rechtlichen und wirtschaftlichen Rahmenbedingungen gestellt und somit die Entwicklung innovativer Anwendungen ermöglicht. Die Studienautoren empfehlen hier Open Data Initiativen zu stärken und mit weiteren Datenquellen – auch Echtzeitinformationen - zu etablieren, um den Wirtschafts- und Wissenschaftsstandort Österreich weiter zu stärken.

Für Unternehmen wie auch Forschungsinstitutionen ist der Kompetenzerwerb im Bereich Big Data essenziell, um die potentielle Information in den Daten für innovative Anwendungen zu extrahieren und zu verwerten. Um diese Kompetenz zu erlangen werden in Kapitel 5.2.4 unterschiedliche Kompetenzerwerbsmodelle vorgeschlagen, die zur Stärkung und Etablierung des Standorts Österreich im Bereich Big Data führen sollen. Eine diesbezüglich ausgearbeitete Maßnahme zur Kompetenzsteigerung in Österreich ist eine stärkere Vernetzung der unterschiedlichen Stakeholder

im Bereich Big Data. Diese ermöglicht einen effizienten Wissensaustausch und das Auffinden von Ansprechpartner für konkrete Herausforderungen in Wissenschaft und Wirtschaft.

Die Ergebnisse der Studie beziehen sich auf die Bereiche Wissenschaft, Wirtschaft und tertiäre Bildung. Im Bereich Wissenschaft kann festgehalten werden, dass Österreich in vielen Teilbereichen von Big Data sehr hohe Kompetenz vorliegt und bei einigen innovativen Forschungsprojekten beteiligt ist. Beispielsweise ist Österreich in dem Bereich High Performance Computing mit wenigen Forschungsgruppen, welche federführend an internationalen Projekten mitwirken, mit hoher Kompetenz ausgestattet. In den Bereichen Analytics sowie Utilization (z.B. Semantische Technologien und Wissensmanagement) verfügt Österreich über mehrere international anerkannte Institutionen und ist die Forschungslandschaft breit aufgestellt. Die Bereiche Big Data Speichersysteme, Big Data Plattformen und Programmiermodelle sind derzeit in dieser Hinsicht in Österreich unterrepräsentiert. Im Bereich Wirtschaft kann als Resultat festgehalten werden, dass der österreichische Markt im Bereich Big Data in den nächsten Jahren von derzeit circa 22,98 Millionen EUR auf 72,85 Millionen Euro steigen wird und somit ein ähnliches Wachstum wie der internationale Markt vorweisen wird können. Österreich verfügt derzeit über einige wenige Firmen mit Know-how in dem Bereich Big Data (siehe Kapitel 2.3) welche nur teilweise am internationalen Markt erfolgreich sind. Auf Grund dessen wird die meiste Wertschöpfung derzeit und in den nächsten Jahren durch nicht-österreichische Unternehmen sowie direkt im Ausland generiert (circa Zwei-Drittel). Dieser Effekt kann durch zielgerichtete Förderungen österreichsicher Unternehmen verringert werden. International wird der österreichische Markt als nicht sehr innovationsfreudig angesehen und die international erfolgreichsten Firmen planen derzeit wenig Präsenz am österreichischen Markt. Zusammengefasst bedeutet dies einen gewissen Aufholbedarf für den Standort Österreich gerade auch in Hinsicht Standortattraktivität und Förderung von innovativen Startups. Sowohl von Forschungsinstitutionen und Unternehmen wurde auch auf die Notwendigkeit von Kompetenzbereitstellung in der tertiären Bildung im Bereich Big Data hingewiesen. Hier wird klar die Etablierung von Data Science Studienplänen auf Basis der vorhandenen Kompetenz und Angebote sowie die Unterstützung von fachspezifischen Weiterbildungsmaßnahmen empfohlen (siehe Kapitel 5.3.7).

Aus Sicht der Forschung und auch der Wirtschaft scheint somit die österreichische Situation in vielen Big Data Bereichen als international nicht federführend. Es bedarf einiger konkreter Schritte, gerade in der Bereitstellung von Kompetenz und in der internationalen Sichtbarkeit, um den Forschungs- und Wirtschaftsstandort für die Zukunft zu stärken und als innovativen und hochqualifizierten Standort zu etablieren. Dieser dargestellten Gesamtsituation folgen auf den derzeitigen Voraussetzungen aufbauende konkrete Empfehlungen für die Stärkung von Big Data in Österreich welche als Überblick in Abbildung 21 dargestellt sind und alle fünf konkreten, in der Studie definierten, Ziele für Österreich verfolgen:

- **Wertschöpfung erhöhen**
- **Wettbewerbsfähigkeit steigern**
- **Sichtbarkeit der Forschung und Wirtschaftsleistung erhöhen**
- **Internationale Attraktivität des Standorts steigern**
- **Kompetenzen weiterentwickeln und festigen**

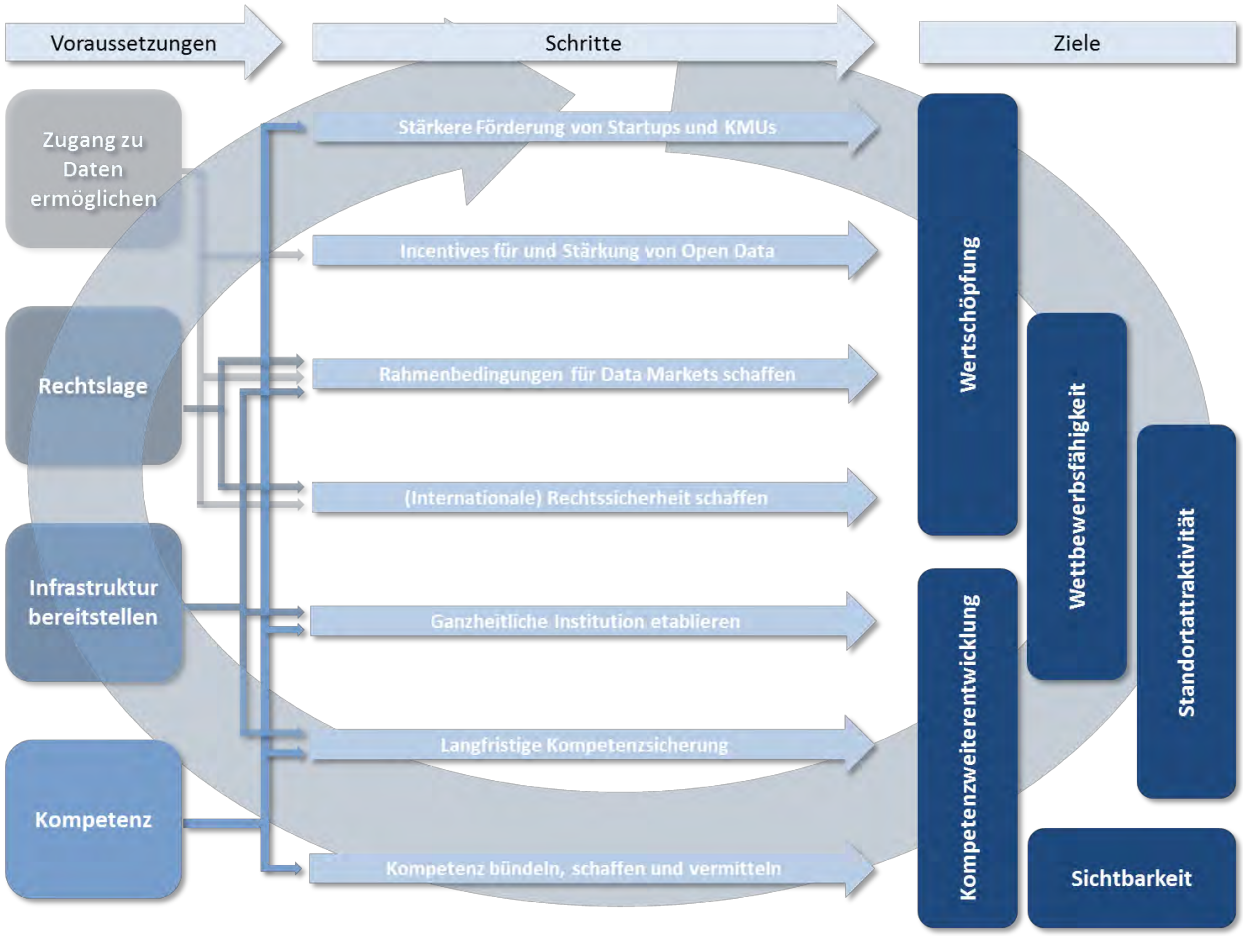


Abbildung 21: Roadmap für Österreich im Bereich Big Data

In Abbildung 21 (Links) sind vier spezifizierte Basisvoraussetzungen für die Durchführung erfolgreicher Big Data Projekte angeführt - Datenzugang, Rechtslage, Infrastruktur und Kompetenz. Die Pfeile die von den Voraussetzungen wegführen, zeigen dabei welche Schritte bzw. Maßnahmen (Mitte) beitragen können, um die jeweiligen Voraussetzungen zu erfüllen und somit die entsprechenden Ziele (Rechts) zu erreichen. Hierbei wurden in der Studiendurchführung folgende Schritte identifiziert:

- **Stärkere Förderung von Startups und KMUs**
- **Incentives für und Stärkung von Open Data**
- **Rahmenbedingungen für Data Markets schaffen**
- **(Internationale) Rechtssicherheit schaffen**
- **Ganzheitliche Institution etablieren**
- **Langfristige Kompetenzsicherung**
- **Kompetenz bündeln, schaffen und vermitteln**

Bei allen Schritten handelt es sich um kontinuierliche Umsetzungen unterschiedlicher Detailmaßnahmen. Dies ist durch den hinterlegten Rundpfeil illustriert. Im Folgenden wird näher auf die einzelnen Ziele, Voraussetzungen und die für die Erreichung der Ziele notwendigen Schritte eingegangen.

5.1 Ziele für eine florierende Big Data Landschaft

In der Durchführung der Studie wurden fünf Kernelemente identifiziert, die zur Stärkung der österreichischen Big Data Landschaft beitragen sollen. Im Folgenden werden diese Ziele näher diskutiert.

5.1.1 Wertschöpfung erhöhen

Ein großes Ziel für den Standort Österreich ist es, neue Wertschöpfung zu generieren und somit das BIP nachhaltig zu stärken, sowie Technologie und Innovationsgetriebene Arbeitsplätze in Österreich zu halten und zu schaffen. Ein wichtiger Faktor hierbei ist internationale Unternehmen im Big Data Bereich am Wirtschaftsstandort Österreich anzusiedeln. Hierbei ist es wichtig, dass diese vor allem stärker in Richtung Forschung und Entwicklung integriert werden und nicht nur vertriebstechnisch mit Niederlassungen vertreten sind. Durch gezielte Maßnahmen in Richtung Einsatz von Big Data Lösungen können vorhandene Unternehmen in Österreich gestärkt werden und sich damit einen Wettbewerbsvorteil erarbeiten. Hierbei geht es primär um die Anwendung von solchen Lösungen. Neben der aktiven Ansiedlung von internationalen Unternehmen und der Förderung der Big Data Anwendung bei heimischen Konzernen ist es auch wichtig, dass Startups und KMUs, welche international erfolgreich sein können, aktiv gefördert werden.

5.1.2 Wettbewerbsfähigkeit steigern

Neben der Generierung von Wertschöpfung ist es wichtig, die Wettbewerbsfähigkeit österreichischer Unternehmen sicherzustellen und zu steigern, um die Entwicklung des Wirtschaftsstandorts Österreich abzusichern. Dies kann auf zwei Ebenen erfolgen:

- Wettbewerbsfähigkeit heimischer Unternehmen: durch die Förderung der Anwendung von Big Data Technologien in nicht-IT Unternehmen können diese Unternehmen ihre Wettbewerbsfähigkeit gegenüber anderen Unternehmen verbessern. Hierbei kann gezielt durch Schulungsmaßnahmen und Fortbildungsschecks unterstützt werden. Eine zielgerichtete Ausbildung zum Data Scientist ist hier auch auf alle Fälle zielführend.
- Wettbewerbsfähigkeit heimischer IT Startups und KMUs: heimische Startups sollen gefördert werden, damit diese international erfolgreich auftreten können und in ihrem Bereich international wettbewerbsfähig sind und die Marktführerschaft erreichen können.

5.1.3 Sichtbarkeit der Forschung und Wirtschaftsleistung erhöhen

International wie auch national ist die Sichtbarkeit eines Standortes und die aktive Kommunikation der vorhandenen Expertise in einem bestimmten Bereich für die Ansiedlung neuer Unternehmen und die Wahrnehmung österreichsicher Marktteilnehmer essenziell. Gerade im Kontext der europäischen Markt und Forschungslandschaft ist es essenziell die Kompetenzen und Stärken eines Marktes wie Österreich hervorzuheben und international zu präsentieren. Ziel hierbei sollte es sein Österreich als Technologie und Innovationsaffinen Markt zu präsentieren. Wirtschaftsförderungsinstitute wie beispielsweise das ZIT (Wirtschaftsagentur Wien) oder AWS können Big Data Leistungen aus Österreich verstärkt in deren Portfolio aufnehmen und diese für internationale Präsentationen verwenden. Durch Netzwerkveranstaltungen von zum Beispiel der OCG können diese auch auf nationaler Ebene für eine bessere Sichtbarkeit sorgen. Eine Dachmarke, welche sich aktiv um das Marketing und der Sichtbaren Gestaltung kümmert, wäre hierbei denkbar. Alternativ kann auch eine bestehende Organisation stärker ausgebaut werden.

5.1.4 Internationale Attraktivität des Standorts steigern

Die Notwendigkeit, auf internationaler Ebene erfolgreich zu sein wurde bereits in den vorangegangenen Punkten erwähnt. Durch einen dedizierten Punkt soll hier noch eine stärkere Betonung darauf gelegt werden. Für die Wettbewerbsfähigkeit Österreichs ist es essenziell, die internationale Bedeutung und Attraktivität zu heben. Gerade in einem technologiegetriebenen Bereich wie Big Data ist hohe Kompetenz am Arbeitsmarkt ein enorm wichtiges Ziel um die Standortattraktivität zu heben. Durch die Hebung der Attraktivität des Standorts können neue Unternehmen, mit Schwerpunkt Forschung & Entwicklung, in Österreich angesiedelt werden, innovative Unternehmen gefördert werden und somit die Wettbewerbsfähigkeit Österreichs erhöht werden.

5.1.5 Kompetenzen weiterentwickeln und festigen

Ein wesentlicher Aspekt für die Steigerung der Attraktivität, der Sichtbarkeit, sowie der Generierung von Wertschöpfung ist die Bereitstellung und Weiterentwicklung von hoher Kompetenz am Arbeitsmarkt. Diese hohe Kompetenz ist in einem innovationsstarken Bereich wie Big Data erforderlich um neue Ideen und Unternehmen zu entwickeln, bestehende zu stärken und internationale Unternehmen anzuziehen. Hierfür wird als wichtiges Ziel die Weiterentwicklung und Festigung der vorhandenen Kompetenzen definiert, welches unter anderem durch eine Stärkung von wissenschaftlich fundierten Ausbildungen und dem Angebot an zusätzlichen hochqualitativen Weiterbildungsmaßnahmen bestehen kann.

5.2 Voraussetzungen und Rahmenbedingungen

In der Folge werden aktuelle Voraussetzungen und Rahmenbedingungen für den Bereich Big Data näher beschrieben. Diese werden als wesentliche Treiber und Hemmnisse für die erfolgreiche Weiterentwicklung des Wirtschafts- und Forschungsstandorts Österreich angesehen. Diese bilden die Grundlage für die zu setzenden Schritte.

5.2.1 Zugang zu Daten ermöglichen

Im Mittelpunkt von Aktivitäten im Bereich Big Data stehen natürlich die Daten selbst. Zur Förderung österreichischer Forschungseinrichtungen und Unternehmen ist der Zugang zu diesen daher ein essentieller Aspekt. Hierbei ist zwischen zwei unterschiedlichen Bereichen und Arten von Datenquellen, extern zur Verfügung gestellten Daten und intern verfügbaren Daten, zu unterscheiden. Immer mehr Unternehmen und Organisationen stehen große Datenmengen zur Verfügung. Um diese Daten intern ausschöpfen zu können werden gerade in Bezug auf personenbezogene Daten klare Richtlinien benötigt (siehe auch Rechtslage - 5.2.2). Ein wesentliches Szenario ist immer öfters die externe zur Verfügung Stellung von Unternehmens- oder Forschungsdaten. Hierbei wird zwischen zwei wesentlichen Vorgehensweisen unterschieden. Einerseits existieren für die Bereitstellung von externen Daten Initiativen in Richtung Open Data und gerade Open Government Data um Daten. Ziel hierbei ist die freie Bereitstellung der Daten um den Markt und auch die eigene Organisation durch innovative Szenarien zu stimulieren. Andererseits können interne Unternehmens- oder Forschungsdaten (zum Beispiel Sensordaten oder Kartenmaterial) auch kommerziell vertrieben werden. In Bezug auf beide Arten der Bereitstellung von Datenquellen muss in der Vorgehensweise sehr stark unterschieden werden. Nichtsdestotrotz gilt es für beide Bereiche klare Richtlinien zur Verfügung zu stellen welche den Zugang zu Daten regeln und dementsprechend auch den Zugang und deren Verwendung ermöglichen.

5.2.2 Rechtslage

Die Datennutzung ist, gerade wenn es sich um (indirekt) personenbezogene Daten handelt, ein sensitiver Bereich der begleitend auch soziale Implikationen auslösen kann (beispielsweise in Bezug auf den Schutz der Privatsphäre). Daher und aufgrund derzeitiger öffentlicher Diskussionen über Datensammlungen (z.B. Vorratsdatenspeicherung) sowie der aktuellen unsicheren nationalen und internationalen Rechtslage, wird dieser Bereich von vielen österreichischen Unternehmen eher gemieden. Eine häufige Aussage dazu ist, dass vor Innovationen im Datenbereich auf Grund der rechtlichen Situation in Österreich zurückgeschreckt wird. Hierbei wurde nicht auf eine strenge Regulierung des österreichischen Marktes verwiesen, sondern auf eine zu unsichere Rechtssituation die Unternehmen ein zu hohes Risiko in Bezug auf Verfahrenssicherheit und die Einschätzung der rechtlichen Situation kalkulieren lässt. Diese Aussagen wurden sowohl von Unternehmen im Big Data Technologiebereich als auch im Anwendungsbereich getätigt. Somit scheint die aktuelle Rechtslage in diesem Bereich eher innovationshemmend zu sein. Als konkrete Zielsetzung wird hier die aktive Implementierung einer international gültigen Rechtslage gesehen. Österreich kann sich hier durch die Etablierung als „Schweiz der Daten“ (Kompa, 2010) einen zusätzlichen Marktvorteil schaffen.

5.2.3 Infrastruktur bereitstellen

Für die erfolgreiche Durchführung von Forschung aber auch für die Entwicklung von Startups und für die Wirtschaft ist die Verfügbarkeit von Infrastruktur (siehe Kapitel 2.2.4) essenziell. Nach derzeitigem Stand steht in Österreich weder auf der Management noch auf der Platform Ebene des Big Data Stacks ein ganzheitlicher Ansatz oder ein einheitlicher Zugriff auf Ressourcen zur Verfügung. Die Entwicklung einer gemeinsamen Strategie aller Kooperationsplattformen und die gezielte Integration und Kommunikation der Anforderungen mit anderen existierenden Initiativen und Plattformen (akademischer aber auch industrieller Natur) in Österreich (ÖAW, ACOnet, EuroCloud, etc.) und auch auf internationaler Ebene findet derzeit nur in eingeschränktem Ausmaß statt. In Österreich findet meistens nur ein projekt- und fachbereichsspezifischer Aufbau von Computing Ressourcen statt (ÖAW, ZAMG, etc.). Themenübergreifende Ansätze (z.B. ACSC und VSC) sind derzeit nicht in dem Maße integriert, wie es für eine gemeinsame Forschungs- und Entwicklungsstrategie notwendig wäre. Im Bereich Cloud Computing, welcher für Big Data eine hohe Wichtigkeit hat, wird für eine gemeinsame Plattform auf das Best Practice Beispiel Okeanos global[46] des griechischen akademischen Netzwerks GRNet auf europäischer Ebene verwiesen werden welcher seit 18. Dezember 2013 zur Verfügung steht.

In Bezug auf den Aufbau der benötigten Infrastruktur wird hier auf die Empfehlungen des europäischen Strategieforums für Forschungsinfrastrukturen (ESFRI) verwiesen. Diese empfiehlt den hierarchischen Aufbau einer europäischen HPC Infrastruktur, welche gerade für Data Science beziehungsweise Big Data genutzt werden kann. Wichtig hierbei ist die Umsetzung nationaler Rechenzentren, welche nicht an bestimmten Universitäten angesiedelt sind sondern als eigenständige Organisationen gebildet werden. Dies bündelt die Infrastruktur-Kompetenzen eines Landes und stellt Kapazitäten für die größten Forschungsprojekte aller nationalen Forschungseinrichtungen (Universitäten, Fachhochschulen, Akademien der Wissenschaften, außeruniversitäre Forschungseinrichtungen) zur Verfügung.

[46] Okeanos global: https://okeanos-global.grnet.gr

5.2.4 Kompetenz in Wirtschaft und Forschung

Wie in Kapitel 2.3 und 0 dargestellt, ist in Österreich in allen Bereichen des Big Data Stacks (siehe Kapitel 2.1.3) und unabhängig von Forschung und Wirtschaft Kompetenz vorhanden. Mit über 20 Fachhochschulen und den österreichischen Universitäten gibt es in Österreich ein breites Angebot an tertiärer Bildung. Derzeit werden von mehreren Fachhochschulen und Universitäten in den angebotenen Masterstudiengängen einige der Themenbereiche des Big Data Stacks abgedeckt. Eine komplette Umsetzung einer Ausbildungsschiene zum Data Scientist konnte von den Studienautoren in Österreich nicht gefunden werden. Die Umsetzung einer kompetenten Ausbildung auf tertiärem Bildungsniveau zum Data Scientist ist aus Sicht des Bereichs Big Data für die Weiterentwicklung des Standorts Österreich essenziell. In Bezug auf die österreichische Forschungslandschaft wird ein sehr breites Spektrum an Kompetenz von vielen unterschiedlichen Institutionen angeboten, welche Teilbereiche des Big Data Stacks abdecken. Eine komplette Abdeckung des Big Data Stacks kann demgegenüber aber nur an wenigen Institutionen gefunden werden.

Zur umfassenden Kompetenzstärkung der österreichischen Big Data Landschaft ist auf eine verstärkte Kooperation von Wissenschaft, tertiärer Bildung und Wirtschaft zu achten. Neben der verstärkten Förderung von anwendungsorientierten Forschung ist hier jedenfalls auch auf den Erhalt der österreichischen Grundlagenforschung zu achten, die bei Innovationsprozessen im Allgemeinen eine wesentliche Rolle spielt und durch die das Basiswissen generiert wird.

5.3 Maßnahmen für die erfolgreiche Erreichung der Ziele

Zur Förderung der österreichischen Big Data Landschaft werden im Folgenden einige Maßnahmen empfohlen, die basierend auf den in Kapitel 5.1 und Kapitel 5.2 besprochenen Zielen und Rahmenbedingungen abgeleitet wurden. Als erste Maßnahme wird die Verteilung und Umsetzung des Leitfadens „Best Practice für Big Data-Projekte“ um die Sichtbarkeit der Möglichkeiten sowie das Bewusstsein für neue Technologien in Österreich zu stärken gesetzt.

5.3.1 Stärkere Förderung von Startups und KMUs

Startups sind ein wichtiger Impulsgeber für die Wirtschaft, für die Steigerung der Wertschöpfung und für die Attraktivierung des Standorts. So finden sich in Österreich einige Förderprogramme zur Gründung neuer Unternehmen. Das daraus resultierende wirtschaftliche Potenzial wird mit den zur Verfügung stehenden Mitteln derzeit aber nur teilweise genutzt und es ist für Firmen sehr schwierig in ihrem Bereich eine gute Marktdurchdringung zu erreichen. Um die Umsetzung nachhaltiger Geschäftsmodelle besser zu unterstützen wird empfohlen den Zugang für Startups zu Venture Capital in Österreich zu vereinfachen. Hierbei sind mehrere Aktionen möglich:

- Verstärkte Unterstützung von Startup-Zentren: Startup-Zentren, die während der Gründungsphase einen Arbeitsplatz zur Verfügung stellen, sollten vermehrt gefördert werden. Diese Förderung kann zeitlich limitiert sein.
- Reduzierung der Gründungskosten für neue Unternehmen: Für die Gründung von Unternehmen sind oftmals höhere Summen notwendig. Eine weitere Senkung dieser Kosten kann den Einstieg in die Selbstständigkeit erleichtern und somit zu neuen Innovationen führen.
- Finanzielle Unterstützung des Unternehmens in der Gründungsphase: Unterschiedlichste Förderinitiativen in Österreich reduzieren derzeit für Personen das finanzielle Risiko bei der

Unternehmensgründung. Allerdings erfordern manche durch deren zeitliche Limitierung kurzfristige Erfolge des Unternehmens. Diese sind oftmals Eingangs eines Innovationsprozesses sehr schwer zu erreichen. Beispielsweise hatte Google erst nach vielen Jahren ein Geschäftsmodell, ähnlich verhielt es sich mit Twitter, Facebook, Amazon und vielen weiteren international erfolgreichen Unternehmen. Eine zeitliche Verlängerung der Förderzeit könnte die Innovationsfähigkeit und das Wachstum eines Unternehmens fördern.

- Etablieren einer „Culture of Failing“ und eines „Soft fails“: In der amerikanischen Kultur ist es erlaubt, mit der eigenen Idee nicht erfolgreich zu sein – weitere Versuche sind erwünscht. Eine dahingehende Bewusstseinsbildung in Österreich kann die Etablierung von international erfolgreichen KMUs weiter fördern.
- Verstärkte Unterstützung von Experten aus Netzwerken: Expertenkreise wie der Arbeitskreis „Cloud und Big Data“ der OCG sowie der Plattform „Digital Networked Data“ besitzen starkes Know-how und kennen die internationalen Möglichkeiten in den jeweiligen Bereichen. Die Förderung von Beratungen für KMUs durch diese Organisationen kann die Erfolgswahrscheinlichkeit nationaler Unternehmen steigern.
- Errichtung einer Venture Capital Plattform: In Österreich sind private Venture Capital Geber nur in einem geringen Ausmaß vorhanden. Es wird empfohlen Maßnahmen für die stärkere Berücksichtigung und Einbindung dieser Investoren zu prüfen und gegebenenfalls durch den Staat zu unterstützen. Denkbar wäre hier die Bereitstellung einer gemeinsamen staatlich geregelten Vermittlungsplattform.

5.3.2 Incentives für und Stärkung von Open Data

Open Data ist ein wesentlicher Aspekt in der Bereitstellung von Daten. Gerade der Bereich von Open Government Data kann hier als Vorzeigerolle dienen und innovationsfördernd wirken. In Österreich werden derzeit einige Datensätze nach dem Open Government Prinzip angeboten (siehe Kapitel 2.5) und dieser Bereich erfährt ein großes Wachstum. Einige Unternehmen, gerade KMUs, und auch viele Forschungsinstitute zeigen das Potenzial dieser Datenquellen auf und implementieren neue innovative Anwendungsfälle (z.B.: (Dax, Transparenz und Innovation durch offene Daten, 2011)). Aus diesem Grund wird klar empfohlen den Bereich Open Data zu stärken und als Innovationstreiber für den Standort Österreich anzusehen.

Um die vollständigen Potenziale von Open Data in Österreich ausschöpfen zu können müssen dennoch einige weitere wichtige Schritte gesetzt werden. Der grundsätzliche Tenor in Gesprächen in Bezug auf Verwendung von Open Data in Unternehmen und Forschung war positiv. Dennoch ist auf mehrere vorherrschende Problematiken hingewiesen worden. Einerseits liegen Daten derzeit in den unterschiedlichsten Formaten und Qualität vor. Dieselben Datensätze werden derzeit häufig von unterschiedlichen Gebietskörperschaften in unterschiedlichen Formaten, Vollständigkeit, Granularität, und unterschiedlichen zeitlichen Auflösungen angeboten (siehe Kapitel 2.5). Weiters ist meistens nicht gewährleistet, dass Daten gewartet werden und in absehbarer Zeit erneuert werden oder weiterhin zur Verfügung stehen werden. Dies erschwert die innovative Nutzung dieser Daten in Unternehmen und der Forschung. In dieser Hinsicht wird die Weiterentwicklung der Open Data Strategie in den Punkten Data Curation und Data Governance empfohlen. Zusätzlich sollte die Zurverfügungstellung von Open Data in Bezug auf Detailgrad, Zeitraum, Formate, Inhalt österreichweit koordiniert werden um die Datenqualität zu erhöhen und langfristig gewährleisten zu können.

Im Bereich Open Data gibt es neben Open Government Data zusätzliche enorme Potenziale die österreichweit genutzt werden können und welche die Umsetzung von neuen Geschäftsideen und die Entwicklung des Standorts maßgeblich positiv beeinflussen können. Als positives Beispiel kann hier die Bereitstellung von Daten der Wiener Linien genannt werden welche in der Bereitstellung von neuen Forschungen und Applikationen resultiert und somit neues Wirtschaftswachstum und Wertschöpfung generiert. Die Öffnung von zusätzlichen Datenquellen im staatsnahen Bereich für die Masse an österreichischen Unternehmen, Forschungseinrichtungen und die große Community der KMUs kann hier für den Wirtschaftsstandort sehr förderlich wirken.

5.3.3 Rahmenbedingungen für Data Markets schaffen

International werden viele Daten von Unternehmen und Organisationen gesammelt und verwertet um einen Marktvorteil zu generieren. Viele Unternehmen haben enorme Datenmengen zur Verfügung welche intern ausgewertet werden können um dem Unternehmen einen Marktvorteil zu verschaffen. Sobald es sich bei Daten um personenbezogene Daten handelt wird eine gesonderte Behandlung in Hinsicht Datenschutz und Sicherheit notwendig. Gerade im Zusammenhang mit internationalen Unternehmen und Organisationen wird diese Herausforderung in Österreich im internationalen Vergleich in den letzten Jahren immer offensichtlicher. Weiters ergibt sich aus der steigenden Datenmenge und den vorhandenen Informationen ein neues Geschäftsfeld des Datenbrokers der aggregierte und angereicherte Datensätze weiterverkauft. Ohne dieses Geschäftsfeld als positiv zu bewerten, ist festzustellen, dass es sich hierbei um einen aufstrebenden Markt handelt und hiermit international enorme Umsätze generiert werden. In der Durchführung dieser Studie war eine häufige Aussage von österreichischen Unternehmen in Bezug auf diese Situation, dass vor Innovationen im Datenbereich auf Grund der rechtlichen Situation in Österreich zurückgeschreckt wird. Dieselbe Aussage gilt für Forschungseinrichtungen, welchen der Zugriff auf notwendiges und relevantes Datenmaterial für konkrete Forschungsfragen oft erschwert wird. Hierbei wurde nicht auf eine strenge Regulierung des österreichischen Marktes verwiesen, sondern auf eine zu unsichere Rechtssituation die Unternehmen vor Investitionen zurückschrecken lässt. Diese Aussagen wurden sowohl von Unternehmen im Big Data Technologiebereich als auch im Anwendungsbereich getätigt (starke Implikationen mit Schritt 5.3.4).

Eine Empfehlung hierbei ist die Förderung der Umsetzung und Etablierung von Marktplätzen für Daten. Mit Marktplätzen können transparente rechtliche Rahmenbedingungen für den Weiterverkauf von Daten geschaffen werden und es kann somit neues Wirtschaftswachstum generiert werden. Hierbei gelten als Ziele Standardverträge zwischen Datenanbietern, Unternehmen welche Daten aufbereiten, und Endkunden welche zusätzliche Informationen in eigene Dienstleistungen und Produkte einbinden können. Die Etablierung von flexiblen Preismodellen kann in diesem Bereich auch förderlich wirken. Als positiver Referenzmarktplatz ist hier die Plattform CloudEO[47] zu nennen welche es Datenanbietern (Satelittendaten) ermöglicht ihre Daten zur Verfügung. Verwertern wird hier ermöglicht diese Daten unter festgelegten Bedingungen zu analysieren und eigene Services auf Basis dieser Daten anzubieten. Durch vorgegebene Preismodelle und feste Aufteilung von Einnahmen zwischen den Entwicklern und den Datenprovidern ist davon auszugehen, dass sich dieses Modell für einen Daten Marktplatz in der Zukunft als Vorzeigemodell für Daten-getriebene Innovationen etablieren kann.

[47] CloudEO AG: http://www.cloudeo-ag.com/

5.3.4 (Internationale) Rechtsicherheit schaffen

Die aktuell vorherrschende Rechtslage in Österreich im Bereich Datenschutz und Datenverarbeitung scheint den Umfragen zufolge in einigen Aspekten innovationshemmend zu sein. Aus diesem Grund fokussiert diese Maßnahme auf die Stärkung des österreichischen Markts durch die Schaffung von Klarheit in datenschutzrechtlichen Thematiken. Hierfür werden drei konkrete Maßnahmen empfohlen:

- **„Schweiz der Daten“:** Als konkrete Umsetzung wird die Etablierung von Österreich als die Schweiz der Daten empfohlen. Dies bedeutet die einfache Umsetzung, und vor allem auch die Durchsetzung, von Datenschutzvorschriften in Österreich und auch das in die Pflicht nehmen von ausländischen Unternehmen. Durch die Etablierung von Österreich als wirklicher „Safe Harbour“ für Daten kann sich Österreich eine weltweite Nische sichern und sich somit einen wesentlichen internationalen Marktvorteil schaffen. Ziel hierbei sollten klare Richtlinien für Datenzugriff und Datenschutz sein sowie deren einfache Umsetzung und Kontrolle. Dieser kann sehr innovationsfördernd wirken und Wertschöpfung generieren. Hierfür ist ein österreichischer Rechtsrahmen mit starkem Fokus und Mitteln für dessen Durchsetzung von Nöten.
- **Internationaler Rechtsrahmen:** Für die Innovationsförderung in Unternehmen und das Entstehen von neuer Wertschöpfung in diesem Bereich muss als klares Ziel ein international gültiger Rechtsrahmen gelten. In einem ersten Schritt sollte Österreich die baldige Umsetzung der geplanten europäischen Richtlinien (Europäische Kommission, 2012) forcieren und in weiterer Folge diese Bemühungen internationalisieren. Die derzeit herrschende Rechtslage in Österreich ist laut Interviews nicht nur innovationshemmend, sondern verschafft den heimischen Unternehmen sogar einen internationalen Nachteil gegenüber Ländern in denen Datenschutz nicht, oder auch klar, geregelt ist. Auf diesen Wettbewerbsnachteil für österreichische Unternehmen sollte in zukünftigen Schritten fokussiert werden um in datengetriebener Wirtschaftsleistung den internationalen Anschluss nicht zu verlieren.
- **Datenschutz und personenbezogene Daten:** Als weiterer Maßnahme (impliziert durch vorhergehende) ist die Umsetzung klarer und vor allem transparenter (internationale) Richtlinien für die Speicherung, die Analyse und auch den Weiterverkauf gerade von (indirekt) personenbezogenen Daten zu sehen. Hierbei ist großer Wert auf den Schutz der Privatsphäre der betroffenen Personen zu legen, jedoch gleichzeitig die Nutzung zu ermöglichen.

5.3.5 Ganzheitliche Institution für Data Science etablieren

Für den nationalen sowie internationalen Erfolg und die Etablierung als technologieaffinen Standort ist die Unterstützung und Sichtbarkeit in den Bereichen neuer Technologien enorm wichtig. Diese Sichtbarkeit kann von einzelnen Marktteilnehmern beziehungsweise Forschungsinstituten nur sehr schwierig erreicht werden, wodurch gewisse Herausforderungen am Markt entstehen. Als Beispiel setzt derzeit Deutschland sehr stark auf Innovation und Kompetenz im Bereich Big Data. Deutschland etabliert derzeit das Smart Data Innovation Lab[48] mit dem Ziel die Voraussetzungen für Spitzenforschung im Bereich Data Engineering/Smart Data für Industrie und Forschung zu verbessern. Derzeit werden im Smart Data Innovation Lab 22 Industriepartner, zwei Verbände sowie zehn unterschiedliche Forschungsinstitutionen vereint. Diese konzentrierte Aufstellung setzt ein

[48] Smart Data Innovation Lab: http://www.sdil.de/de/

klares internationales Zeichen in Richtung Forschungs- und Industrieführerschaft in dem Bereich Big Data und generiert enorme internationale Sichtbarkeit. Zusätzlich wird derzeit auch das Berlin Big Data Center (BBDC)[49] eingerichtet welches die Forschungsaktivitäten von mehreren Universitäten vereint und verstärkt. Mit diesem Hintergrund zeigt sich, dass Sichtbarkeit und strategische Positionierung in neuen und innovativen Forschungs- und Wirtschaftszweigen enormes Potenzial und Wichtigkeit für die Entwicklung und Sichtbarkeit des Marktes und Wirtschaftsstandort besitzt. Um dies für Österreich ebenfalls zu erreichen und den heimischen Wirtschaftsstandort als innovationsfreudig zu präsentieren, empfehlen die Studienautoren ebenfalls eine gemeinsame Plattform oder Institution für Data Science zu etablieren. Wichtig hierbei ist es, die vielfältigen bestehenden Kompetenzen die in Österreich vorhanden sind zu vereinen und gemeinschaftlich als international kompetitiver Standort aufzutreten. Eine Data Science Institution sollte als Anlaufstelle für Forschung, Wirtschaft und gerade auch für KMUs dienen und allen eine Kooperation und internationale Sichtbarkeit unter einer Dachmarke ermöglichen. Bei dieser Maßnahme ist auf zu erwartenden hohen Aufwand hinzuweisen, welcher sich aber auch in enormen Potenzial und einer sehr stark verbesserten Innovationskraft widerspiegeln wird.

Ein wichtiger Aspekt in Bezug auf die Umsetzung einer Data Science Institution für Österreich ist die Kompetenzaneignung für die Wirtschaft. Kompetenz ist der Schlüssel für die effiziente Umsetzung von Big Data Projekten und deren erfolgreiche Abwicklung. Eine Big Data Plattform sollte es der Wirtschaft ermöglichen gezielt Kompetenzen im Big Data Umfeld zu erwerben und zu vertiefen. Hierbei sind zwei Schritte notwendig:

- **Spezifische Ausbildung und Weiterbildungsangebote im Big Data Bereich**: Einerseits besteht eine Notwendigkeit für Kompetenzentwicklung in der tertiären Bildung in Richtung Big Data spezifischer Technologien. Hierbei wird die Weiterentwicklung einer fundierten Grundlagenausbildung sowohl im theoretischen als auch im praktischen Bereich empfohlen. Gerade in sich schnell weiterentwickelten Bereichen wie Big Data ist es notwendig eine fundierte theoretische Ausbildung bereitzustellen, um sowohl in der Forschung als auch in der Wirtschaft die weiteren Entwicklungen maßgeblich beeinflussen und diese in die verschiedenen Anwendungsdomänen erfolgreich integrieren zu können. Ein Ziel hierbei kann eine Flexibilisierung des vorhandenen Angebots sein um vorhandene Kompetenzen auf neue Art und Weise für innovative Ausbildungsmodelle zu bündeln. Des Weiteren besteht hohes Potenzial für konkrete Weiterbildungsmaßnahmen für Unternehmen um Mitarbeiter an neue Technologien heranzuführen um in weiterer Folge innovationsführend wirken zu können.
- **Erfahrungen und Kompetenzerwerb durch die Abwicklung von Big Data Projekten:** Für Unternehmen und wissenschaftliche Organisationen besteht der Bedarf Kompetenz in innovativen Bereichen zu erwerben und diese anhand von durchgeführten Projekten zu vermitteln um in neuen Technologiebereichen Fuß fassen zu können. Bei dieser Eintrittsschwelle kann die Umsetzung einer Institution für Big Data auf mehrere Arten hilfreich sein. Erstens kann eine Institution für Big Data durch die Bündelung von Kompetenz mehrerer Stakeholder als direkter Ansprechpartner für internationale und nationale Kooperationen dienen und auf diese Art und Weise sowohl den Wirtschafts- als auch Forschungsstandort durch gezielte neue Projektumsetzungen stärken. Dies kann in weiterer Folge zu einer erhöhten internationalen Sichtbarkeit und einer gestärkten Außenwirkung für den Standort Österreich führen. Weiters kann eine Institution für Big Data für noch nicht in

[49] Berlin Big Data Center (BBDC): http://www.pressestelle.tu-berlin.de/medieninformationen/2014/maerz_2014/medieninformation_nr_432014/

diesem Marktsegment vertretene Unternehmen und Forschungseinrichtungen als erster Ansprechpartner für neue Projekte und Kooperationen dienen und auf diese Art und Weise den Kompetenzerwerb in diesen Organisationen unterstützen. Hierbei wird ein Modell empfohlen, das neuen Teilnehmern die einfache Umsetzung von Erstprojekten ermöglicht und auf diese Art und Weise einen Markteinstieg durch erste Referenzen ermöglicht. Dieser Punkt wurde wiederholt von österreichischen Unternehmen als besonders wichtig hervorgehoben.

Die erfolgreiche Umsetzung eines Kompetenzzentrums für Big Data begründet sich auf dem gemeinsamen Angehen aller Ebenen des Big Data Stacks. Die Autoren dieser Studie empfehlen Aktivitäten im Bereich Infrastruktur, Plattformen, Analyse und Utilization zu bündeln um das volle Potenzial für Österreich ausschöpfen zu können. Im Bereich Infrastruktur fehlt Österreich eine hierarchisch aufgebaute und integrierte Infrastruktur welche in einem kompetitiven Umfeld von Unternehmen und Forschungseinrichtungen verwendet werden kann. Hier wird empfohlen, dass die Bereitstellung von Infrastruktur mit einer (Big Data) Institution gekoppelt wird um die gesamte Kompetenz zu bündeln. Hierbei sind folgende Punkte zu beachten:

- **Eigenständigkeit:** Die Organisation sollte nicht Teil einer existierende Universität oder Forschungseinrichtung sein, sondern eine eigenständige Rechtskörperschaft bilden oder eine bestehende Netzwerkorganisation ergänzen. Auf diese Art und Weise kann die Einbindung neuer Forschungseinrichtungen und Unternehmen auf einfache und klar definierte Art und Weise erfolgen.
- **Steuerung durch alle teilnehmenden Organisationen:** Die teilnehmenden Forschungseinrichtungen sowie Unternehmen müssen Steuermöglichkeiten in Bezug auf die Weiterentwicklung und den Zugriff auf diese Organisation haben um eigene Anforderungen in die Institution einbringen zu können.
- **Balance aus Forschung und Betrieb:** Die Weiterentwicklung und auch der Betrieb der Organisation müssen forschungsgetrieben sein, um Innovationen über den State-of-the-Art bereitstellen zu können. Hierbei ist auf die Integration der Bedürfnisse von Wirtschaft und Forschung zu achten und es muss ein Betrieb auf hohem technischen Niveau sichergestellt werden um innovative Infrastrukturen und Plattformen zur Verfügung stellen zu können.
- **Nationale Kontaktstelle für europäische Plattformen und Institutionen:** Diese Organisation sollte Österreich auf europäischer Ebene im Bereich Big Data strategisch vertreten und die Koordination mit internationalen Institutionen zu fördern.
- **Offenheit:** Die Institution sollte offen sein und allen österreichischen Forschungseinrichtungen und Unternehmen zur Verfügung stehen.
- **Nachhaltigkeit**: Für die erfolgreiche Umsetzung einer Institution in diesem Bereich ist eine nachhaltige Finanzierung abseits von Projektfinanzierungen notwendig.
- **Agilität:** Ziel der Institution soll die nachhaltige Umsetzung und Bereitstellung von innovativen Technologien im Bereich Big Data und deren Integration in verschiedene Anwendungsdomänen sein. Hierfür ist eine sehr hohe Agilität und Adaptivität zu neuen Entwicklungen und Anwendungsfeldern notwendig.

Eine erfolgreiche Umsetzung einer Big Data Institution bedarf der direkten Einbindung unterschiedlicher Anwendungsfelder und wichtiger Stakeholder in Forschung und Wirtschaft(z.B. OCG Arbeitsgruppe, Digital Networked Data, Open Data Austria). Das Smart Data Innovation Lab in Deutschland konzentriert sich dabei auf vier wichtige Anwendungsfelder: Industrie 4.0, Energie, Smart Cities und Medizin. Für Österreich wird eine flexiblere Einbindung und Integration von Anwendungsgebieten empfohlen. Als wichtige in der Studie klassifizierte Anwendungsfelder wird die

Miteinbeziehung der Bereiche Verkehr, Gesundheitswesen, Energie, und Raumfahrt und Industrie 4.0 empfohlen.

5.3.6 Langfristige Kompetenzsicherung

Die Stärkung von Kompetenz und Wissen ist für die Weiterentwicklung eines Standorts enorm wichtig. Gerade in hochtechnologischen Bereichen ist das Vorhandensein von hochqualitativer Expertise die Basis für das Entstehen von Innovationen aus der Forschung und der Wirtschaft - und somit von Wertschöpfung. Dazu ist eine hochqualitative tertiäre Bildungslandschaft notwendig, die fundierte grundlagenorientierte Ausbildung anbietet, hochaktuelle Themen aus der Forschung miteinbezieht und praxisrelevante Themenbereiche behandelt. Dies erfordert breite Expertise in den unterschiedlichsten Forschungsbereichen und Anwendungen. Hierzu ist eine langfristige Sicherung der Grundfinanzierung von Bildung und Forschung & Entwicklung notwendig.

Für Spitzenforschung und die daraus resultierenden Innovationen und deren Wertschöpfung wird der langfristige Aufbau von Kompetenzen benötigt. Dazu werden in der Forschungslandschaft entsprechende Finanzierungen und Planstellen benötigt. Der anhaltende Trend zu kurzfristigen Stellenbesetzungen an Österreichs Universitäten erschwert die Aufrechterhaltung und Sicherung der vorhandenen Expertise in Österreich und fördert den Brain Drain in andere Länder. Die anwendungsorientierte Forschung wird in Österreich stark durch thematische Programmlinien gefördert und gesteuert. Um hier nachhaltig Kompetenz zu sichern, ist einerseits die langfrsitige Finanzierung einzelner Programmlinien sicherzustellen. Andererseits sollten die thematischen Ausrichtungen der Forschungsprogramme in den thematischen Vorgaben den individuellen Innovationsspielraum nicht einschränken. In dieser Hinsicht wird für die Entwicklung des Big Data Standorts Österreich empfohlen, die vorhandenen Kompetenzen im Bereich Big Data (Grundlagen und anwendungsorientierte Forschung) durch langfristige Finanzierungen abzusichern und auszubauen und explizit zu fördern.

5.3.7 Kompetenz bündeln, schaffen und vermitteln

Grundsätzlich besteht eine Notwendigkeit für Kompetenzentwicklung in der tertiären Bildung in Richtung Big Data spezifischer Technologien. Hierbei wird die Etablierung einer fundierten Grundlagenausbildung sowohl im theoretischen als auch im praktischen Bereich empfohlen. Gerade in sich schnell weiterentwickelten Bereichen wie Big Data ist es notwendig eine fundierte theoretische Ausbildung bereitzustellen und zu erweitern um sowohl in der Forschung als auch in der Wirtschaft nicht nur den aktuellen Stand der Entwicklungen verstehen zu können als auch die weiteren Entwicklungen maßgeblich beeinflussen zu können. Die Kombination mit anwendungsspezifischen und praxisorientierten Themengebieten ist hier jedenfalls wünschenswert.

Ein Ziel hierbei kann eine Flexibilisierung des vorhandenen Angebots sein um vorhandene Kompetenzen auf neue Art und Weise für neue Ausbildungsmodelle zu bündeln. Einige tertiäre Bildungseinrichtungen decken den gesamten Big Data Stack ab und sind in der Lage ein vollständiges Data Science Curriculum anzubieten. Eine bessere Sichtbarkeit dieser Kompetenzen und eine flexible Integration der verteilten Teilaspekte in neuen Masterstudiengängen, oder generell offenen Masterstudien (z.B. über Kernfachkombinationen), kann eine einfache aber sehr wirksame Möglichkeit bieten. Weiters könnten durch eine Flexibilisierung Kompetenzen von unterschiedlichen Einrichtungen in diesem Bereich sehr effizient gebündelt werden. Als Beispiel wird hier ein Joint-Curriculum in Kombination einer technischen Fakultät und einer wirtschaftswissenschaftlichen Fakultät in Richtung spezialisierter Data Analysts genannt.

Zusätzlich besteht hohes Potenzial für konkrete Weiterbildungsmaßnahmen für Unternehmen um Mitarbeiter an neue Technologien heranzuführen und in weiterer Folge innovationsführend wirken zu können. In Kombination von tertiären Bildungseinrichtungen oder Forschungseinrichtungen mit Unternehmen können maßgeschneiderte Weiterbildungsangebote im Bereich Big Data entwickelt und angeboten werden. Eine Förderung in diesem Bereich könnte hier ein großes Potenzial für hochqualitative Weiterbildungsmaßnahmen für Unternehmen bieten.

Den Umfragen zufolge wird der Europäische Markt von Unternehmen noch nicht im potentiellen Ausmaß wahrgenommen - viele Unternehmen sehen den heimischen Markt als einzig relevanten an. Um das Potential des Europäischen Markts besser auszuschöpfen, wird von den Studienautoren empfohlen, die europäische Idee stärker zu vermitteln. Dadurch kann die Wettbewerbsfähigkeit gesichert und gestärkt werden, um in weiterer Folge zusätzliche Wertschöpfung zu generieren.

6 Literaturverzeichnis

AIT Mobility Department. (2013). *Strategy 2014-2017.* Vienna.

al., D. V. (2014). *IDC's Worldwide Storage and Big Data Taxonomy, 2014.* Framingham: IDC.

al., D. V. (2012). *Perspective: Big Data, Big Opportunities in 2013.* Framingham: IDC.

al., D. V. (2013). *Worldwide Big Data Technology and Services 2013–2017 Forecast.* Framingham: IDC.

Alpaydin, E. (2010). *Introduction to Machine Learning, second edition.* London, England: The MIT Press.

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks, 54(15)* , S. 2787-2805.

Banerjee, P., & et. al. (2011). Everything as a service: Powering the new information economy. *Computer 44.3* , S. 36-43.

Battre, D., Ewen, S., Hueske, F., Kao, O., Markl, V., & Warneke, D. (2010). Nephele/PACTs: A Programming Model and Execution Framework for Web-Scale Analytical Processing. *Proceedings of the ACM Symposium on Cloud Computing (SoCC)* .

Bendiek, S. (2014). *itdaily*. Abgerufen am 08. April 2014 von http://www.it-daily.net/analysen/8808-die-zukunft-gehoert-den-data-scientists

Berners-Lee, T. (2001). The semantic web. *Scientific American* , S. 28-37.

Bierig, R., Piroi, F., Lupu, M., Hanburry, A., Berger, H., Dittenbach, M., et al. (2013). Conquering Data: The State of Play in Intelligent Data Analytics.

Big Data Public Private Forum. (2013). *Consolidated Technical Whitepapers.*

Big Data Public Private Forum. (2013). *First Draft of Sector's Requisites.*

BITKOM. (2012). *Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte.*

BITKOM. (2013). *Leitfaden: Management von Big-Data-Projekten.*

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. In *International Journal on Semantic Web and Information Systems (IJSWIS) 5.3* (S. 1-22).

BMVIT. (2014). *Bundesministerium für Verkehr, Innovation und Technologie* .

Booch, G., Rumbaugh, J., & Jacobson, I. (1998). The Unified Modeling Language (UML). *http://www. rational. com/uml/(UML Resource Center)* .

Brewer, E. (02 2012). Pushing the CAP: Strategies for Consistency and Availability. *IEEE Computer, vol. 45, nr. 2* .

Brewer, E. (2000). Towards Robust Distributed Systems. *Proceedings of 9th Ann. ACM Symp. on Principles of Distributed Computing (PODC 00)* .

Brown, B., Chui, M., & Manyika, J. (2011). *Are you ready for the era of 'big data'?* http://unm2020.unm.edu/knowledgebase/technology-2020/14-are-you-ready-for-the-era-of-big-data-mckinsey-quarterly-11-10.pdf: McKinseyQuarterly.

Bundeskanzleramt. (2014). *Digitales Österreich*. Abgerufen am April 2014 von http://www.digitales.oesterreich.gv.at/site/5218/default.aspx

Buyya, R., Broberg, J., & Goscinski, A. (2011). *Cloud Computing: Principles and Paradigms.* Wiley.

Cheng T. Chu, Sang K. Kim, Yi A. Lin, Yuanyuan Yu,, & Gary R. Bradski, Andrew Y. Ng, Kunle Olukotun. (2006). Map-Reduce for Machine Learning on Multicore. *NIPS* .

Cisco. (2014). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018.*

Cloudera. (kein Datum). *Cloudera Impala, Open Source, Interactive SQL for Hadoop*. Abgerufen am 2 2014 von http://www.cloudera.com/content/cloudera/en/products-and-services/cdh/impala.html

Colling, D., Britton, D., Gordon, J., Lloyd, S., Doyle, A., Gronbech, P., et al. (01 2013). Processing LHC data in the UK. *Philisophical transactions of the royal society* .

Cooperation OGD Österreich: Arbeitsgruppe Metadaten. (2013). *OGD Metadaten - 2.2.* Abgerufen am April 2014 von https://www.ref.gv.at/uploads/media/OGD-Metadaten_2_2_2013_12_01.pdf

Critchlow, T., & Kleese Van Dam, K. (2013). *Data-Intensive Science.* CRC Press.

Davenport, T., & Patil, D. (2013). *Harvard Business Review*. Von http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ abgerufen

Dax, P. (2011). Transparenz und Innovation durch offene Daten. *futurezone.at* .

Dax, P., & Ledinger, R. (November 2012). *Open Data kann Vertrauen in Politik stärken*. Von http://futurezone.at/netzpolitik/11813-open-data-kann-vertrauen-in-politik-staerken.php abgerufen

Dean, J., & Ghemawat, S. (2008). Mapreduce: simplified data processing on large clusters. *Communications of the ACM, vol. 51* , S. 107-113.

(2013). *Demystifying Big Data - A Practical Guide To Transforming The Business of Government.* TechAmericaFoundation.

Dowd, K., Severance, C., & Loukides, M. (1998). *High performance computing. Vol.2.* O'Reilly.

DuBois, L. (2013). *Worldwide Storage in Big Data 2013–2017 Forecast.* Framingham: IDC.

Edlich, S., Friedland, A., Hampe, J., Brauer, B., & Brückner, M. (2011). *NoSQL Einstieg in die Welt nichtrelationaler Datenbanken.* Hanser.

Eibl, G., Höchtl, J., Lutz, B., Parycek, P., Pawel, S., & Pirker, H. (2012). *Open Government Data – 1.1.0.*

Ekanayake, J., Li, H., Zhang, B., Gunarathne, T., & Bae, S.-H. (2010). Twister: A Runtime for Iterative MapReduce. *19th ACM International Symposium on High Performance Distributed Computing* .

Ekanayake, J., Pallickara, S., & Fox, G. (2008). MapReduce for Data Intensive Scientific Applications. *IEEE International Conference of eScience* .

EMC2. (2014). Abgerufen am 08. April 2014 von https://infocus.emc.com/david_dietrich/a-data-scientist-view-of-the-world-or-the-world-is-your-petri-dish/

EU Project e-CODEX. (2013). Von http://www.e-codex.eu/ abgerufen

EU Project Envision. (2013). Von http://www.envision-project.eu/. abgerufen

EU Project PURSUIT. (2013). Von http://www.fp7-pursuit.eu/PursuitWeb/. abgerufen

EuroCloud. (2011). *Leitfaden: Cloud Computing: Recht, Datenschutz & Compliance.*

Europäische Kommission. (25. 1 2012). *Vorschlag für VERORDNUNG DES EUROPÄISCHEN PARLAMENTS UND DES RATES*. Von http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0011:FIN:DE:PDF abgerufen

European Commission. (2010). *The European eGovernment Action Plan 2011-2015.*

Fayyad, U., Piatestsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Maganzine* .

Federal Chancellery of Austria. (2013). *What is eGovernment?* Von http://oesterreich.gv.at/site/6878/default.aspx abgerufen

Forbes. (2013). *A Very Short History Of Big Data*. Abgerufen am 16. 08 2013 von http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/

Fraunhofer. (2014). Abgerufen am 08. April 2014 von http://www.iais.fraunhofer.de/data-scientist.html

George, L. (2011). *HBase: The Definitive Guide.* O'Reilly.

Gorton, I., & Gracio, D. (2012). *Data-Intensive Computing: A Challenge for the 21st Century.* Cambridge.

Grad, B., & Bergin, T. J. (10 2009). History of Database Management Systems. *IEEE Annals of the History of Computing Volume 31, Number 4* , S. 3-5.

Grothe, M., & Schäffer, U. (2012). *Business Intelligence.* John Wiley & Sons.

Gu, Y., & Grossman, R. (2009). Sector and Sphere: the design and implementation of a high-performance data cloud. *Philosophical Transactions of the royal society* .

Gu, Y., & Grossman, R. (2007). UDT: UDP-based data transfer for high-speed wide area networks. *Computer Networks* .

Habala, O., Seleng, M., Tran, V., Hluchy, L., Kremler, M., & Gera, M. (2010). Distributed Data Integration and Mining Using ADMIRE Technology. *Grid and Cloud Computing and its Applications, vol. 11 no. 2* .

Haerder, T., & Reuter, A. (1983). Principles of transaction-oriented database recovery. *ACM Computing Surveys, Vol. 15, Nr. 4* , S. 287-317.

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques, 2nd Edition.* The Morgan Kaufmann Series in Data Management Systems.

Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space (1st edition), Synthesis Lectures on the Semantic Web: Theory and Technology.* Morgan & Claypool.

Heise, A., Rheinländer, A., Leich, M., Leser, U., & Naumann, F. (2012). Meteor/Sopremo: An Extensible Query Language and Operator Model. *BigData Workshop (2012)* .

Hewitt, E. (2010). *Cassandra: The Definitive Guide.* O'Reilly.

Hey, T., & Trefethen, E. (2005). Cyberinfrastructures for e-Science. Science. *Science Magazine* .

Hey, T., Tansley, S., & Tolle, K. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. *Microsoft Research* .

HL7. (2013). *HL/ Electronic Health Record*. Von http://www.hl7.org/ehr/. abgerufen

Hüber, B., Kurnikowski, A., Müller, S., & Pozar, S. (2013). *Die wirtschaftliche und politische Dimension von Open Government Data in Österreich.*

IBM. (2011). *IBM Big Data Success Stories.*

IBM. (2014). *What is a data scientist?* Von http://www-01.ibm.com/software/data/infosphere/data-scientist/ abgerufen

IEEE. (1984). *Guide to Software Requirements Specification. ANSI/IEEE Std 830-1984 .* Piscataway/New Jersey: IEEE Press.

IHT SDO. (2013). *SNOMED CT*. Von http://www.ihtsdo.org/snomed-ct/ abgerufen

IT Cluster Wien. (2012). *Software as a Service – Verträge richtig abschließen 2., erweiterte Auflage.*

Jackson, M., Antonioletti, M., Dobrzelecki, B., & Chue Hong, N. (2011). Distributed data management with OGSA-DAI. *Grid and Cloud Database Management* .

Johnston, W. (1998). High-Speed, Wide Area, Data Intensive Computing: A Ten Year . *Seventh IEEE International Symposium on High Performance Distributed Computing* .

KDNUggets. (2014). *KDNUggets*. Abgerufen am 08. April 2014 von http://www.kdnuggets.com

Keim, D., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). Visual analytics: Scope and challenges. In *Visual Data Mining* (S. 76-90). Berlin: Springer.

Kell, D. (2009). In T. Hey, S. Tansley, & K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery.*

Khoshafian, S., Copeland, G., Jagodits, T., Boral, H., & Valduriez, P. (1987). *A Query Processing Strategy for the Decomposed Storage Model.* ICDE.

Kidman, A. (2014). Abgerufen am 08. April 2014 von http://bigdataanalyticsnews.com/choose-best-tool-big-data-project/

Koehler, M. (2012). A service-oriented framework for scientific Cloud Computing.

Koehler, M. e. (10 2012). The VPH-Share Data Management Platform: Enabling Collaborative Data Management for the Virtual Physiological Human Community. *The 8th International Conference on Semantics, Knowledge & Grids, Beijing, China* .

Koehler, M., & Benkner, S. (2009). A Service Oriented Approach for Distributed Data Mediation on the Grid. *Eighth International Conference on Grid and Cooperative Computing* .

Koehler, M., Kaniovskyi, Y., Benkner, S., Egelhofer, V., & Weckwerth, W. (2011). A cloud framework for high troughput biological data processing. *International Symposium on Grids and Clouds, PoS(ISGC 2011 & OGF 31)069* .

Kompa, M. (2010). Wird Island die Schweiz der Daten? *Telepolis* .

Kreikebaum, H., Gilbert, D. U., & Behnam, M. (2011). *Strategisches Management.* Stuttgart: Kohlhammer.

Leavitt, N. (14. 02 2010). Will NoSQL Databases Live Up to Their Promise? *IEEE Computer,vol.43, no.2* .

Malewicz, G., Austern, M., Bik, A., Dehnert, J., Horn, I., Leiser, N., et al. (2010). Pregel: a system for large-scale graph processing. *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* .

Markl, V. (28. November 2012). *Big Data Analytics – Technologien, Anwendungen, Chancen und Herausforderungen.* Wien.

Markl, V., Löser, A., Hoeren, T., Krcmar, H., Hemsen, H., Schermann, M., et al. (2013). *Innovationspotentialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen (Big Data Management).* BMWi.

Mason, H. (2014). Abgerufen am 08. April 2014 von http://www.forbes.com/sites/danwoods/2012/03/08/hilary-mason-what-is-a-data-scientist

McGuiness, D., & Van Harmelen, F. (2004). OWL web ontology language overview. *W3C recommendation* .

McKinsey&Company. (04 2013). *How big data can revolutionize pharmaceutical R&D*. Von http://www.mckinsey.com/insights/health_systems_and_services/how_big_data_can_revolutionize_pharmaceutical_r_and_d abgerufen

Melnik, S., Gubarev, A., Long, J., Romer, G., Shivakumar, S., Tolton, M., et al. (2010). Dremel: Interactive Analysis of Web-Scale Datasets. *Proceedings of the VLDB Endowment* .

Message Passing Interface Forum. (21. September 2012). MPI: A Message-Passing Interface Standard.

Moniruzzaman, A. M., & Akhter Hossain, S. (2013). NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison. *ArXiv e-prints* .

NESSI. (2012). *Big Data - A New World of Opportunities.*

Open Government Data. (2013). Abgerufen am 2013 von http://gov.opendata.at

OpenMP Application Program Interface, Version 4.0. (July 2013). OpenMP Architecture Review Board.

Organisation of American States. (2013). *e-Government*. Von http://portal.oas.org/Portal/Sector/SAP/DptodeModernizaci%C3%B3ndelEstadoyGobernabilidad/NPA/SobreProgramadeeGobierno/tabid/811/language/en-US/default.aspx abgerufen

Poole, D., & Mackworth, A. (2010). *Artificial Intelligence, Foundations of Computational Agents.* Cambridge University Press.

Popper, K. (2010). *Die beiden Grundprobleme der Erkenntnistheorie: aufgrund von Manuskripten aus den Jahren 1930-1933.* Vol. 2. Mohr Siebeck.

Proceedings of ESWC. (2013).

Sabol, V. (2013). Visual Analytics. TU Graz.

Sendler, U. (2013). *Industrie 4.0–Beherrschung der industriellen Komplexität mit SysLM (Systems Lifecycle Management).* Springer Berlin Heidelberg.

Shim, J., Warketin, M., Courtney, J., Power, D., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems, Volume 33, Issue 2* .

Tachmazidis, I. et al. (2012). Scalable Nonmonotonic Reasoning over RDF data using. MapReduce. *Joint Workshop on Scalable and High-Performance Semantic Web Systems (SSWS+ HPCSW 2012).*

Tiwari, S. (2011). *Professional NoSQL.* wrox.

Top500 Supercomputer Sites. (08 2013). Von http://www.top500.org/. abgerufen

Trillitzsch, U. (2004). Die Einführung von Wissensmanagement. *Dissertation* . St. Gallen, Schweiz.

Ussama, F., & et. al. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. American Association for Artificial Intelligence .

Valiant, L. (1990). A bridging model for parallel computation. *Communications of the ACM, vol.33, no. 8* , S. 103-111.

W3C. (2013). *OWL Web Ontology Language*. Von http://www.w3.org/TR/webont-req/#onto-def. abgerufen

Wadenstein, M. (2008). *LHC Data Flow.*

Wahlster, W. (1977). *Die repräsentation von vagem wissen in natürlichsprachlichen systemen der künstichen intelligenz.* Universität Hamburg, Institut für Informatik.

Warneke, D., & Kao, O. (2009). Nephele: Efficient Parallel Data Processing in the Cloud. *Proceedings of the 2nd Workshop on Many-Task Computing on Grids and Supercomputers* .

Wirtschaftskammer Österreich. (2013). *Die österreichische Verkehrswirtschaft, Daten und Fakten - Ausgabe 2013.*

Wrobel, S. (2014). Big Data – Vorsprung durch Wissen. *BITKOM Big Data Summit.*

Yen, W. (2014). *Using Big Data to Advance Your Cloud Service and Device Solutions*. Abgerufen am 08. April 2014 von http://channel9.msdn.com/Events/Build/2014/2-645

Zhang, Y., Meng, L., Li, H., Woehrer, A., & Brezany, P. (2011). WS-DAI-DM: An Interface Specification for Data Mining in Grid Environments. *Journal of Software, Vol 6, No 6* .

Anhang A Startworkshop #BigData in #Austria

Der Startworkshop zu #BigData in #Austria wurde am Donnerstag den 12. Dezember 2013 von 9:30 bis 14:00 im MEDIA TOWER, Taborstrasse 1-3, 10 Wien im World Café Stil durchgeführt. Die Ergebnisse der Diskussionen der circa 40 TeilnehmerInnen von Unternehmen und Universitäten sind direkt in die Studie mit eingeflossen. Hier sind die Themengebiete, die externen DiskussionsleiterInnen sowie die konkreten Fragestellungen gelistet.

Big Business - Marktchancen und Use Cases/Leitdomäne

Diskussionsleitung: Thomas Gregg, Braintribe IT Technologies GmbH

- Wie ist das Feedback Ihrer KundInnen bzw. in Ihrem Unternehmen? Wie wird Big Data wahrgenommen?
- In welchen Brachen/Domänen sehen Sie Potenziale für Big Data Technolgien? (z.b. Gesundheitswesen weil …, Handel und Verkauf weil …, Verkehrswesen weil …)
- Welche Domänen profitieren von dem Einsatz von Big Data?
- Welche Projekte kennen Sie, wo Big Data eingesetzt wurde/wird?
- Welche Branchen/Domänen weigern sich, Big Data Lösungen einzusetzen?
- Wie wird Big Data in Ihrem Unternehmen aktuell eingesetzt?
- Welche Potenziale sehen Sie durch einen Einsatz von Big Data Technologien?
- Welche Domänen können als Vorreiter für Big Data Technologien am österreichischen Markt fungieren?
- Wie kann Österreich im Bereich Big Data eine Führungsrolle einnehmen?
- Welche Potenziale sehen Sie gerade für österreichische Unternehmen um durch den Einsatz von Big Data zu profitieren?

Big Challenges - Herausforderungen in und Anforderungen an Big Data Projekte

Diskussionsleitung: Gerhard Ebinger, Teradata

- Welche Probleme gibt/gab es in der Umsetzung von Big Data Projekten?
- Welche Erfahrungen wurden mit Big Data Projekten soweit gemacht?
- Welche Schritte sind durchzuführen, um ein Big Data Projekt aufzusetzen?
- Was sind domänenspezifische Anforderungen für den erfolgreichen Einsatz von Big Data-Verfahren?
- Wie Unterscheiden sich Big Data Projekte von anderen Projekten?
- Welche Verfahren, Plattformen und Technologien erleichtern die Umsetzung von Big Data Projekten?
- Welche spezifischen Herausforderungen sehen Sie in Ihrer Institution in Bezug auf Big Data (Technische Anforderungen, Verfügbarkeit und Verwendung von Daten, Big Data Projekte)?
- Welche spezifischen Anforderungen sehen Sie in Bezug auf die Umsetzung und Planung von Big Data Projekte in Ihrem Unternehmen? (Technische Anforderungen, Personelle Anforderungen, Organisatorische Anforderungen, Daten-spezifische Anforderungen)
- Welche Verfahren und Werkzeuge im Bereich Big Data sehen Sie als große Herausforderung (Recheninfrastruktur, Skalierbare Speicherlösungen, Datenmanagement und Integration, Datenanalyse Software, Datenrepräsentation (Visualisierung, Wissensbasierte Systeme))?

Big Science - Wissenschaftliche Entwicklungen und Perspektiven im Bereich Big Data)

Diskussionsleitung: ao. Univ.-Prof. Peter Brezany, Universität Wien, Fakultät für Informatik

- Wie ist die wissenschaftliche Akzeptanz und Definition von Big Data (Stichwort: fourth paradigm)?
- Was sind die wichtigsten Forschungsfragen in der Umsetzung von Big Data Technologien?
- Welche Potenziale, Anforderungen und Risiken gibt es in der anwendungsorientierten Big Data Forschung?
- Stichwort “Data Scientists”: Potenziale in der Ausbildung, Forschung und am Arbeitsmarkt?
- Wie ist die österreichische Forschungslandschaft in Bezug auf Big Data aufgestellt?
- Wie kann der österreichische Standort durch die Förderung von Big Data-spezifischen Themen profitieren?

Big Datenschutz - Datenschutz, Governance, rechtliche Aspekte und Auswirkungen

Diskussionsleitung: Dipl.-Ing. Mag Walter Hötzendorfer, Universität Wien, Fakultät für Rechtswissenschaft

- Welche (rechtlichen) Rahmenbedingungen gibt es für die Verwendung von öffentlichen Datenquellen?
- Welche rechtlichen Einschränkungen gibt es für Weiterverarbeitung von Daten (Unternehmensintern/extern)?
- Was sind rechtliche und Datenschutzspezifische Anforderungen für den Einsatz von Big Data-Verfahren?
- Welche rechtliche Einschränkungen und Möglichkeiten gibt es für den Verkauf und Ankauf von Daten?
- Welche Auswirkungen haben Überwachungstechnologien auf die Datenverwendung in Ihrer Organisation?
- Welche Einschränkungen sind beim Zugriff auf Daten zu beachten?
- Wie werden soziale und gesellschaftspolitische Auswirkungen bei der Verarbeitung von personenbezogenen Daten berücksichtigt?
- Wie sehen Sie die Relevanz und Akzeptanz von Datenschutz in Ihrer Organisation?

Anhang B Interviews Anforderungen, Potenziale und Projekte

Die Interviews wurden auf unterschiedliche Arten durchgeführt:

- Online: hierbei wurde mit IDC-Tools eine Umfrage zur Akzeptanz von Big Data in Österreich durchgeführt. Bei dieser Art der Umfragen bestehen qualitative Mechanismen, welche sicherstellen, dass die Qualität der Daten höchsten Ansprüchen genügt. Befragt wurden Unternehmen aus Österreich, welche in den verschiedensten Industrien tätig sind (ausgenommen direkter IT Unternehmen). Letztere wurden dediziert ausgeschlossen, da diese nicht als KäuferInnen von Lösungen im Big Data Umfeld in Frage kommen. Dabei handelt es sich um eine „Demand-Side" Umfrage. Das Sample hierbei war n>150.
- Interviews: Die Supply-Side wurde durch direkte Gespräche mit Big Data Verantwortlichen der jeweiligen Unternehmen durchgeführt. Hierbei liegt der Fokus auf offene Gespräche, wo auf die Fragen der Marktakzeptanz, der aktuellen und erwarteten Umsätze sowie der Strategie gestellt wurden.

Anhang C Disseminationworkshop #BigData in #Austria

Im Rahmen der Studie #BigData in #Austria wurden zwei Dissemniation Workshops abgehalten. Der erste Dissemination Workshop wurde am 11.04.2014 in Link in der Arche Noah, Wirtschaftskammer Oberösterreich mit 15 TeilnehmerInnen abgehalten. Der zweite Dissemination Workshop wurde in Wien am 16.4.2014 in Zusammenarbeit mit der OCG in der Wollzeile 1 mit 45 TeilnehmerInnen abgehalten. Im Rahmen der Disseminationworkshops wurden die Studienergebnisse präsentiert und im Anschluss diskutiert. Die Ergebnisse der Diskussionen wurden in der Folge in die Studienresultate eingearbeitet.