

# ***Potenzialeinschätzung von Big Data Mining als methodischer Zugang für Foresight***

Christian K. Bosse, Judith Hoffmann, Ludger van Elst

## ***Zusammenfassung***

In den letzten Jahren haben sich rund um den Begriff „Big Data“, durch die im Zuge der digitalen Transformation erheblich gesteigerte Menge an (oft unstrukturierten) Informationen und die Möglichkeiten zur Auswertung hochvolumiger Datenbestände, hohe gesellschaftliche Erwartungen entwickelt. Auch im Bereich Foresight wird über den Einsatz von Big Data-Technologie diskutiert, insbesondere seitdem erste Assistenzsysteme und Dienstleistungsangebote am Markt verfügbar sind. Inwiefern diese jedoch den Anforderungen im Kontext von Foresight entsprechen und welche Potenziale im Technologieeinsatz stecken, wurde bislang nicht ausreichend untersucht. Aus diesem Grund wurde im Rahmen eines Forschungsprojektes eine Potenzialanalyse durchgeführt, die Anforderungen, Möglichkeiten und Grenzen von Big Data-Technologie für Foresight-Prozesse untersucht, sie den aktuell bestehenden Angeboten gegenüberstellt und daraus resultierende Forschungsbedarfe für zukünftige Arbeiten ableitet.

## ***Abstract***

In recent years, the concept of "big data" has become quite a hype as new possibilities and opportunities in the field of high-volume data set analysis are debated. Also, the use of big data technology is discussed in the context of foresight, especially since first assistance systems and services have been available on the market. However, neither the extent to which these meet the requirements of foresight nor the capabilities connected with the use of big data technology in regard to foresight have so far been analysed sufficiently. Therefore an analysis was performed that aimed to examine the requirements, capabilities and limitations of big data technology in foresight processes. Existing solutions are discussed and areas for further research derived.

## **1. Hintergrund und Ziel der Studie**

Der Begriff „Big Data“ hat sich in den letzten Jahren zu einem Phänomen mit hoher Eigendynamik entwickelt, ohne dass es ein einheitliches Verständnis oder gar eine allgemein gültige Definition gibt. Insbesondere im

Zusammenhang mit Digitalisierung werden Möglichkeiten und Grenzen von „Big Data“ diskutiert (Buhl et al. 2013; Fasel, Meier 2016; Gadatsch 2017). Im weitesten Sinne steht der Begriff „Big Data“ für eine sehr große, exponentiell stetig anwachsende Menge heterogener, meist nur schwach strukturierter Daten aus verschiedenen Quellen.

*“Big Data refers to both large volumes of data with high level of complexity and the analytical methods applied to them which require more advanced techniques and technologies in order to derive meaningful information and insights in real time” (HM Government Horizon Scanning Programme 2014).*

Allein die Frage, ab wann eine Datenmenge als „Big Data“ bezeichnet und mit manuellen bzw. klassischen Methoden der Datenverarbeitung nicht mehr ausgewertet werden kann, führt zu verschiedenen Einschätzungen. Darüber hinaus wird der Begriff synonym für unterschiedliche Analyse- und Auswertungsverfahren von Massendaten verwendet, sodass nur selten eindeutig verständlich ist, was exakt gemeint ist (Bachmann, Kemper, Gerzer 2014). Im Rahmen der vorliegenden Arbeit beschreibt der Begriff den methodischen Zugang zur Sammlung und Auswertung von Datenmengen, die bzgl. Volumen und Komplexitätsgrad so groß sind, dass sie mit herkömmlichen Werkzeugen nicht mehr erfasst und ausgewertet werden können.

Auch in der strategischen Vorausschau, dem sog. Foresight, verspricht der Einsatz von Big Data-Technologie tieferegreifende Erkenntnisse (z.B. Kayser, Blind 2017; Zhang et al. 2016). Sie soll als informationstechnische Basis für die Speicherung und Verarbeitung entsprechend hochvolumiger Datenmengen dienen, die kontinuierlich anfallen und anwachsen, z. B. Publikationsdatenbanken, Patentdaten, soziale Medien oder auch (Selbst)Überwachungsdaten. Die Hoffnung liegt darin, mit Hilfe von Big Data-Technologien aus Massendaten schnell und effizient Informationen zu (neuartigen) Phänomenen und Entwicklungen zu generieren (z.B. McAfee, Brynjolfsson 2012; Rohrbeck et al. 2015). In Organisationen soll dies dazu beitragen, zukünftige Herausforderungen, insbesondere in Zusammenhang mit der digitalen Transformation zu meistern. Nahezu alle Bereiche einer Organisation werden von der Digitalisierung erfasst, so auch das strategische Management (Müller-Seitz et al. 2016) und das Innovationsmanagement (Müller-Seitz, Reger 2010). Dies bringt Änderungen von Arbeits-, Organisations- und Marktstrukturen mit sich und es ist die Frage zu klären, wie mit den damit verbundenen Prozessen und Strategien angemessen umgegangen werden kann (Lingnau et al. 2018). Schließlich soll Foresight als Werkzeug für Entscheidungsfindung und Komplexitätsreduktion dienen (Schatzmann et al. 2013). Dabei ist es eine Herausforderung, mit dem sogenannten „Information overload“ umzugehen, dem Phänomen, dass die Güte von Managemententscheidungen durch ein Zuviel an Informationen nicht nur nicht steigt, sondern sinkt. Bereits in den 1960/70er Jahren erkannte der Wirtschaftsnobelpreisträger Herbert Simon, dass zu viel Information die Aufmerksamkeit der Empfänger negativ beeinflusst (Simon 1971). In einer

Welt des Zuviel werde es zum „ökonomischen Privileg von Reichtum“, vermutet auch der Chicagoer Soziologe Andrew Abbott, dass „Menschen seltener von schlechten Informationen überflutet werden“ (Abbott 2014).

Zum Umgang mit dieser Herausforderung wurden auch für Informationen mit Relevanz für Forschung, Technologie und Innovation Ansätze und Lösungen entwickelt, in denen generische Big Data-Basiswerkzeuge angepasst und für spezifische Analyse-, Auswertungs- und Darstellungsschritte zum Einsatz kommen. Jedoch ist in den meisten Fällen intransparent, welche Funktionen die am Markt angebotenen Lösungen enthalten, wie der Prozess der Informationsextraktion abläuft und wie valide die gewonnenen Ergebnisse sind.

Um hier einen generellen Überblick zu erarbeiten und die Möglichkeiten eines Einsatzes von Big Data-Technologie im Kontext der strategischen Vorausschau auszuloten, wurde das Forschungsvorhaben „Potenzialeinschätzung von Big Data Mining als methodischer Zugang für Foresight und Wissenschaftskommunikation“ durchgeführt. Das Institut für Technologie und Arbeit übernahm dabei in Kooperation mit dem Deutschen Forschungszentrum für Künstliche Intelligenz im Rahmen des vom Bundesministerium für Bildung und Forschung geförderten Projektes die Potenzialanalyse im Bereich Foresight. Dieser Teilbereich des Verbundprojektes zielte darauf ab, die konkreten Potenziale und aktuell (noch) bestehenden Herausforderungen des Einsatzes von Big Data-Technologie zu analysieren, bestehende Anforderungen zu erheben sowie die am Markt verfügbaren Systeme zu betrachten. Ferner wurden die gewonnenen Erkenntnisse dazu herangezogen, um zukünftige Forschungsbedarfe zu identifizieren und hervorzuheben.

## **2. Grundlage der Analyse und Ableitung von Anforderungen an die Technologie**

Nachfolgende Kapitel verdeutlichen die Grundlagen der Potenzialanalyse in Form eines Foresight-Einsatzszenarios mit Anwendungsfällen für Big Data-Assistenzsysteme. Auf dieser Basis lassen sich geeignete Informationsquellen eingrenzen und Anforderungen an Assistenzsysteme ableiten.

### **2.1 Anwendungsfälle als Grundlage der Potenzialanalyse**

Der Potenzialanalyse liegt ein Einsatzszenario zugrunde, das allgemein auf breit angelegte Foresight-Prozesse abzielt, wie sie beispielsweise im Kontext der Forschungs-, Technologie- und Innovationspolitik vorzufinden sind. Durch einen dort notwendigen breiten Suchraum, der sowohl technologische als auch gesellschaftliche Entwicklungen umfasst, ergeben sich besonders hohe Anforderungen an einen Quellen- und Methodenmix.

Um möglichst aussagekräftige Ergebnisse zu erhalten, wurde dieses Einsatzszenario mit Hilfe von drei Anwendungsfällen konkretisiert. Diese beschreiben detaillierte Aufgabenstellungen, Einsatzmöglichkeiten und Herausforderungen für einen realitätsnahen Einsatz von Big Data-Technologien im Bereich der strategischen Vorausschau. Die Anwendungsfälle zeigen beispielhaft, dass sowohl die Nutzung unterschiedlicher Informationsquellen, als auch die Anwendung unterschiedlicher methodischer Zugänge je nach Foresight-Fragestellung notwendig sein können. Sie lassen sich wie folgt kurz skizzieren:

- 1) Zentrales Anliegen bei der Beobachtung von Themenfeldentwicklungen und Akteuren ist die Generierung von Orientierungswissen. So soll abgeleitet aus den dynamischen Entwicklungsverläufen von Themenfeldern und inhaltlichen Aspekten ein Überblick über den Entwicklungsstand vieler verschiedener Themen und den damit assoziierten Akteuren generiert werden. Darüber hinaus sollen thematische Schnittmengen, z. B. Verknüpfungspunkte zwischen Themenfeldern, Abhängigkeiten und Wechselwirkungen beschrieben werden. Ziel ist, neben einem Monitoring bekannter Themen, auch potenzielle Synergieeffekte bislang unverbundener Themen bzw. Technologien zu identifizieren und adressierbar zu machen.
- 2) Mit Hilfe eines systematischen und IT-unterstützten Analyseprozesses mit kreativen Elementen sollen Themen außerhalb des Mainstreams und neuartige Entwicklungen anhand schwacher Signale (Vgl. Ansoff (1976): Diskontinuitäten (Ereignisse, die zu empfindlichen Anpassungsmaßnahmen zwingen) können prinzipiell antizipiert werden und kündigen sich durch sogenannte schwache Signale an, S. 131) sowie Schnittstellenthemen, exotische Themen und Lücken mit Zukunftspotenzial identifiziert werden. Speziell solche unerwarteten Entwicklungen bergen ein großes Disruptionspotenzial, das – frühzeitig erkannt – gestaltend adressiert werden kann.
- 3) Anknüpfend an den ersten Anwendungsfall birgt eine themenfeldspezifische Vertiefung die Möglichkeit, Adressaten gezielt mit auf ihre Informationsbedarfe zugeschnittenen Ergebnissen zu versorgen. Aufbauend auf den übergreifenden Erkenntnissen zur Entwicklung der Themenslandschaft, sollen vertiefende Analysen zu Entwicklungen von Inhalten, deren mögliche (Weiter-) Entwicklung und Akteurskonstellationen geliefert werden. Damit wird das Ableiten von Optionen für themenfeldspezifische Strategien oder Maßnahmen unterstützt.

Die skizzierten Anwendungsfälle, wurden u. a. in Form von Anwendergeschichten weiter konkretisiert. Damit wurden sie „greifbar“. Im nächsten Schritt konnten fallbeispielhaft entsprechende Informationsquellen recherchiert werden, um deren Verfügbarkeit und Zugänglichkeit auszuloten. Danach konnten für jeden Anwendungsfall Anforderungen an ein Big-Data-Assistenzsystem abgeleitet werden. Damit war ein eine konkrete Vorstellung über Gestaltungsanforderungen an ein beispielhaftes, idealtypisches Assistenzsystem expliziert. Dies bildete die Grundlage für einen Abgleich mit Informationen zu bestehenden Ansätzen und Systemen. Eine weitere Perspektive bilden Experteneinschätzungen, die mit Hilfe einer qualitativen Erhebung in der Foresight-Fachcommunity erarbeitet wurden.

## 2.2 Informationsbedarfe und -quellen

Der Zugriff auf eine Vielzahl heterogener Informationsquellen erscheint notwendig, um die definierten Anwendungsfälle für Big Data-Systeme im Kontext der strategischen Vorausschau in der Praxis realisieren zu können. Dabei unterscheiden sich die Informationsquellen sowohl im zu verarbeitenden Datenvolumen als auch in ihrer Komplexität bzw. Heterogenität. Ebenso variiert die Absicherung der Datenqualität zwischen den Quellen sehr stark. Je nach Komplexität der Fragestellung müssen verschiedene Informationsquellen gleichzeitig herangezogen und sinnvoll in eine Analyse einbezogen werden, um verlässliche Ergebnisse zur Beantwortung einer Fragestellung zu erhalten. Abbildung 1 zeigt eine grobe Anordnung von Informationsquellen, die für die oben skizzierten Anwendungsfälle nützlich sein könnten, hinsichtlich ihrer Datenvolumina, also der Anzahl und der Größe ihrer Datensätze. Ein Datensatz kann dabei beispielsweise ein Zeitungsartikel, ein Fachartikel oder ein Blogbeitrag sein, der im Volltext oder auch nur als Quellenverweis mit seinen Meta-Daten (Titel, Autor, etc.) vorliegen kann.

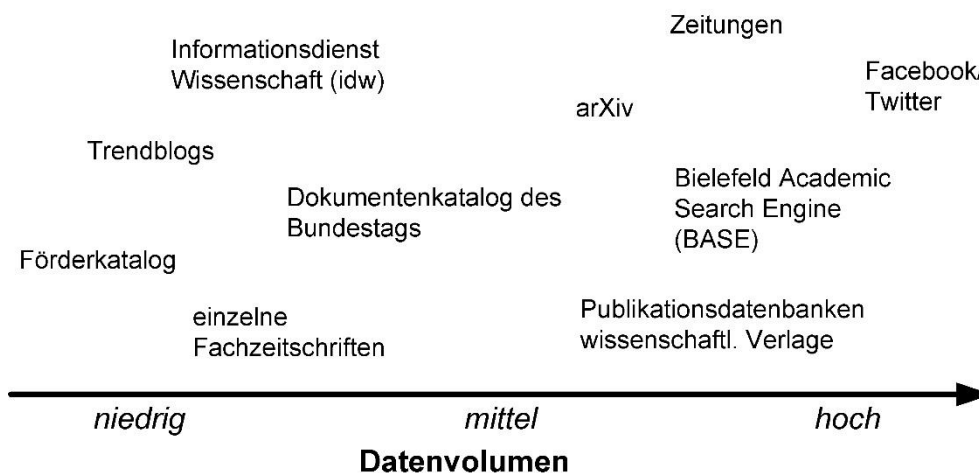


Abbildung 1: Spektrum an beispielhaften Informationsquellen mit ihren unterschiedlichen Datenvolumina (eigene Darstellung)

Teilweise kann auf sehr strukturierte Daten, etwa von Literaturdatenbanken, zurückgegriffen werden, die eine verhältnismäßig einfache Auswertung ermöglichen. In anderen Fällen handelt es sich um unstrukturierte Quellen, bei denen sich die Datenauswertung komplizierter gestaltet, etwa in Kommentaren von Social Media-Angeboten. Für den gewünschten Erkenntnisgewinn können jedoch beide Arten gleichermaßen essenziell sein. Rechtliche Rahmenbedingungen, finanzielle Abwägungen oder Probleme, verschiedene Quellen sinnvoll zu verknüpfen, machen die Auswertung und die dazu notwendigen Vorüberlegungen aufwändig.

Folglich ist die Definition und Zusammenstellung eines einmal gestalteten, allgemeingültigen Quellenportfolios nicht zielführend oder gar unmöglich.

Insbesondere für längerfristig angelegte Auswertungen gilt es, die initial ausgewählten Informationsquellen in regelmäßigen Zyklen zu überprüfen, inaktive Quellen ggf. herauszunehmen sowie neue zu ergänzen. Dies zeigt sich beispielsweise in den Anwendungsfällen 1 und 3, in denen eine kontinuierliche Analyse von Themenfeldentwicklungen im wissenschaftlichen Bereich im Fokus steht. Hierzu wurde ein Portfolio an Informationsquellen – u. a. bestehend aus Online-Datenbanken mit Hochschulschriften, Fachzeitschriften, Informationen zu Forschungsprojekten oder Pressemitteilungen im Bereich der Wissenschaft – zusammengestellt, mit deren Hilfe die aktuellen Stände der Wissenschaft in den zentralen Themenfeldern abgebildet werden können. Im Anwendungsfall 2 hingegen werden sowohl der wissenschaftliche als auch der politische und gesellschaftliche Diskurs kontinuierlich untersucht sowie inhaltlich gegeneinander abgeglichen. Dies bedingt jeweils eine breite Quellenlage zur Abdeckung der unterschiedlichen Perspektiven, die zur Informationsgewinnung herangezogen und kontinuierlich überprüft bzw. erweitert werden kann. So wird bspw. ergänzend zu den wissenschaftlichen Datenbanken der politische Diskurs über die Veröffentlichungen des Deutschen Bundestages abgebildet, oder der gesellschaftliche Diskurs über Soziale Medienplattformen und Medienberichte.

Insbesondere sozialen Medien erscheinen als Informationsquelle in Zusammenhang mit Foresight interessant (Danner-Schröder 2018). Soziale Medien bringen einige Merkmale mit sich, die sie stark von anderen Informationstechnologien unterscheiden. Eine solche Möglichkeit ist, neben der Aktualität, die öffentliche Darstellung eines Netzwerkes. Beziehungen in Kontaktlisten kann eingesehen werden, wer mit wem befreundet ist, auf Pinnwänden lässt sich nachverfolgen, wer mit wem Informationen austauscht und die Beiträge anderer kommentiert (Leonardi, Vaast 2017). Heftige Diskurse zeigen Risiken bei der Nutzung sozialer Medien: „Fake News“, d. h. falsche oder gefälschte Meldungen, und Möglichkeiten diese zu erkennen (siehe bspw. Elkasrawi et al. 2016) und „Filter Bubbles“, d. h. Filterblasen übereinstimmender Informationen, die dadurch entstehen, dass Webseiten mit Algorithmen vorauszusagen versuchen, welche Informationen der Benutzer auffinden möchte (Vgl. Pariser 2012, von Gehlen 2011), sind dabei zentrale Themen. Deren Einfluss sollte bei der Einschätzung der Nützlichkeit von Daten im Rahmen der Nutzung für Foresight-Prozesse berücksichtigt werden. Gleichzeitig verdeutlichen diese Diskurse auch, dass Facebook, Twitter und soziale Medien allgemein zu einem zentralen Anlaufpunkt für aktuelle Berichterstattung geworden sind. Die Nutzer vertrauen den angezeigten Informationen und geben in Form von Likes, Kommentaren und Shares ihre Haltung zu bestimmten Themen preis. Eine gezielte Auswertung dieser Reaktionen wiederum kann die Erstellung einer sehr realitätsnahen Abbildung des gesellschaftlichen Meinungsbilds ermöglichen, wie es zuvor ohne groß angelegte repräsentative Umfragen nicht möglich war. Des Weiteren können ergänzende Informationen oder thematische Verknüpfungen zu den bestehenden Themen zusammengetragen werden, die eine hohe inhaltliche Bereicherung sein können. Und mit einem nach vorn gerichteten Blick bieten Analysen und Auswertungen der Kommentare und Posts in den

sozialen Medien u. a. sogar das Potenzial, neuartige und potenziell zukünftig relevante Themen und gesellschaftliche Trends zu identifizieren.

Bei der Auswahl der Quellen ist neben thematisch passenden Inhalten auch die Art der Quelle zu beachten. Zum einen ist hier zwischen lizenz- bzw. kostenpflichtigen und den frei zugänglichen (Open Access) Quellen, zum anderen zwischen strukturierten Datenbanken und Online-Katalogen sowie unstrukturierten Informationen von Webseiten oder aus den sozialen Medien zu unterscheiden. Letzteres ist dahingehend wichtig, dass unstrukturierte Daten meist erst von Webseiten ausgelesen und anschließend noch händisch strukturiert werden müssen, bevor sie für Analysen und Auswertungen herangezogen werden können. Da dies in den meisten Fällen ein erheblicher Kosten- und Zeitfaktor ist, sollte die jeweilige Möglichkeit der Informationsextraktion bei der Auswahl der passenden Quellen berücksichtigt werden. Bei den sozialen Medien gibt es hierfür spezialisierte Dienstleistungsanbieter, welche die Inhalte themenspezifisch auswerten und die Ergebnisse für weitere Analysen zur Verfügung stellen. Einen Vorteil bieten hingegen die bereits vorab strukturierten Informationsquellen wie zum Beispiel Online-Kataloge von Bibliotheken, Datenbanken von Verlagen oder ähnlichen. Auf diese kann oftmals über bereitgestellte Schnittstellen zugegriffen werden, sodass Anfragen direkt auf der Datenbasis ausgeführt und abgearbeitet werden können. Im Vergleich zu den unstrukturierten Quellen ist diese Direktanbindung der Informationsquelle in den meisten Fällen jedoch kostenpflichtig. Je nach Umfang und Komplexität der Auswertungsanfragen können schnell Kosten von mehr als 100.000 Euro anfallen. Darüber hinaus ergibt sich zumeist bei unstrukturierten Informationsquellen wie sozialen Medien oder Blogs, aber auch bei strukturierten, für Autoren frei zugänglichen Publikationsdatenbanken, eine weitere, höchst relevante Problematik. Während Publikationen in Datenbanken von großen Verlagen oder Bibliotheken strengen Kriterien und einer Begutachtung unterliegen, können in Open Access Datenbanken oder Pre-Printservern Autoren ohne jegliche Begutachtung ihre Texte veröffentlichen. Dies führt dazu, dass vor einer Auswertung die analysierten Inhalte qualitativ überprüft und validiert werden müssen, um vertrauenswürdige Ergebnisse zu erzeugen. Zum Beispiel decken deutsch- und englischsprachige Literaturdatenbanken nicht die weltweite Forschungscommunity ab, oder zu einigen Technologien zeigt sich ein andersartiges Veröffentlichungs- und Patentierungsverhalten als zu anderen.

Eine weitere, generell für das Text- und Data-Mining bestehende Herausforderung ist der Umgang mit dem deutschen Urheberrecht. Die neuen Verfahren des Text- und Daten-Mining, die im Zuge von Big Data-Auswertungen zum Einsatz kommen, unterliegen starken rechtlichen Beschränkungen. Eine Studie des Wellcome Trust aus dem Jahr 2011 verdeutlichte bspw. die Schwierigkeiten, die sich aus der unsicheren Rechtslage ergeben können. Ein Wissenschaftler, der etwa Text Mining in Texten zum Thema „Malaria“ durchführen wollte, hätte demnach ein Jahr mit der reinen Abklärung der benötigten Rechte verbracht und 62 Prozent seines Jahresgehalts für Lizenzzahlungen aufbringen müssen (Wellcome Trust 2012). Dieses Beispiel

zeigt, welchen Herausforderungen Wissenschaftler im europäischen Raum - bzw. konkret in Deutschland - gegenüberstehen, während in den USA, Japan und Großbritannien die Nutzung solcher Verfahren im nicht-kommerziellen wissenschaftlichen Kontext genehmigungsfrei möglich ist.

### **2.3 Anforderungen an ein idealtypisches System**

Die Anforderungen an ein softwaretechnisches Assistenzsystem für den Einsatz von Big Data-Technologie erscheinen in den Anwendungsfällen grundsätzlich sehr ähnlich. Die Kernaufgabe lässt sich grob mit der Identifikation von Mustern sowie dem Bilden von Verknüpfungen und Vernetzungen von einzelnen Dokumenten in einem Datenbestand umschreiben. Diese sollen u. a. dazu eingesetzt werden, anhand bekannter Muster und aktueller Daten zukünftige Entwicklungen oder potenziell relevante Themen herauszuarbeiten. Die hierfür notwendigen Funktionen und Applikationen, wie z. B. Suchfunktionen, Ähnlichkeitsanalysen, die Erfassung von Metadaten oder die Strukturierung von Daten bspw. in Clustern, sind übergreifend einsetzbar und bestehen größtenteils schon als bekannte Routinen. Eine größere Herausforderung auf der technischen Seite ist vielmehr die Integration neuer heterogener Informationsquellen in das System, die möglichst flexibel und leicht handhabbar gestaltet werden muss. Im besten Fall sollte dies sogar derart konzipiert sein, dass das Portfolio an eingebundenen Informationsquellen möglichst durch den Endanwender selbst bearbeitet und erweitert werden kann. Hierbei sind Fragen der Datenqualitätssicherung und der Gewichtung unterschiedlicher Informationsquellen sowie der mit den Quellen repräsentierten Suchfelder konkret zu beantworten.

Weitere Anforderungen werden darüber hinaus an das System an sich gestellt. Insbesondere die Nachvollziehbarkeit und Transparenz der Prozesse und Ergebnisse ist von enormer Relevanz. Die Kernaufgabe der Technologie ist die Unterstützung des Anwenders bei der explorativen Erkundung von schriftlich dokumentierten Informationen aus verschiedenen Quellen. Um die Entstehung der Ergebnisse jederzeit nachvollziehen und die generierten Ergebnisse reproduzieren zu können, wäre ein als „White-Box“ angelegtes System wünschenswert. Das bedeutet, alle Schritte der Ergebnisentstehung sind parallel zu den Arbeiten mit dem Assistenzsystem nachvollziehbar dokumentiert und offengelegt. Neben der Transparenz der Prozesse muss auch das Assistenzsystem als Ganzes den Anforderungen des Anwenders gerecht werden. Die Aufgaben, die mit dem Assistenzsystem unterstützt werden sollen, sind durchweg komplexer Natur, sodass die Bedienbarkeit des Systems ebenfalls nicht trivial ist. Eine einfache, intuitive Anwendung wäre zwar wünschenswert, würde gleichzeitig aber die Einsatzmöglichkeiten und Funktionen zu sehr einschränken. Daher muss ein Mittelweg gefunden werden, der die Anwendung durch einen Experten vorsieht, aber dennoch eine übersichtliche Bedienung und insbesondere übersichtliche Visualisierung der Informationen auf der Benutzeroberfläche garantiert.

Gleichermaßen resultieren Anforderungen an den Menschen, der als „regelmäßiger Anwender“ das Assistenzsystem einsetzt. Hierzu zählt u. a. eine



grundlegende Affinität im Umgang mit neuer Technologie. Ein wesentliches Verständnis für Funktionsweise, abgedeckte Inhalte und Interpretationsrahmen automatisiert generierter Ergebnisse erscheint ebenso als Grundvoraussetzung zur erfolgreichen und sachgerechten Anwendung. Zur explorativen Erkundung sowie zur Analyse und Nachverfolgung von Themenentwicklungen ist bei der Arbeit mit einem Assistenzsystem außerdem das Ausschalten von Wahrnehmungsfiltren zur Ermöglichung der Detektion von Neuem notwendig. Gleichzeitig muss der Anwender über die notwendigen Erfahrungen sowie fachlich-inhaltliche Vorkenntnisse verfügen, um mit dem System ein effektives, exploratives Vorgehen zu erzielen. Ohne diese Erfahrungen, Vorkenntnisse und eine theorie- bzw. hypothesengeleitete Suchstrategie würde das Arbeiten mit dem System einer „orientierungslosen Suche nach der Nadel im Heuhaufen gleichen“, also bestenfalls zu Zufallsfunden führen.

Bei den Untersuchungen wurden darüber hinaus Anforderungen an die prozessuale Einbindung eines solchen Assistenzsystems mit Big Data-Technologie erhoben. So ist beispielsweise eine kontinuierliche Nutzung des Systems durch den Anwender oder ein Team von Anwendern eine Grundvoraussetzung dafür, dass Anwendungsexpertise und -routine entwickelt, aufrechterhalten sowie kontinuierlich erweitert werden können. Des Weiteren muss eine sinnvolle Einbindung des Assistenzsystems in bestehende oder neu zu entwickelnde Prozesse gewährleistet sein, um die generierten Ergebnisse effektiv weiterverarbeiten und nutzen zu können.

Zusammenfassend kann geschlussfolgert werden, dass im Prinzip Big Data-Anwendungen den Anforderungen an sozialwissenschaftliche Methoden (Scheiermann, Zweck 2014), an Methoden der strategischen Vorausschau sowie an digitale Unterstützungssysteme genügen müssen. Eine plakativ-kritische Sicht auf die Einhaltung solcher Anforderungen an Big Data-Anwendungen in der derzeitigen Praxis (Abbildung 2) mahnt zur Aufmerksamkeit bei der Entwicklung und Anwendung entsprechender Systeme.

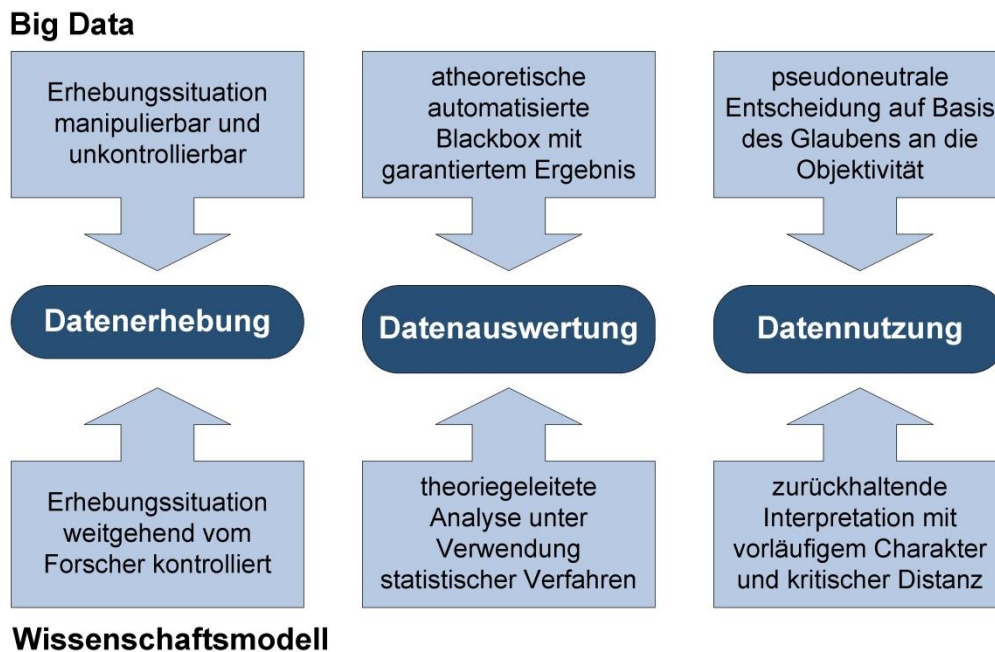


Abbildung 2: Kritisch-pointierte Gegenüberstellung von Big Data und Wissenschaftsmodell (Meyer 2017)

Diese kritisch-pointierte Gegenüberstellung macht grundsätzlich das Problem der Intransparenz komplizierter oder komplexer Algorithmen deutlich und kann auch als Aufforderung zur Entwicklung von Verfahren zur Qualitätssicherung im Sozialwissenschaftlichen Sinne bzgl. Reliabilität und Validität verstanden werden.

### 3. Analyse verfügbarer Systeme

Neben einer vertiefenden Recherche zu bestehenden Lösungen am Markt wurden im weiteren Verlauf der Potenzialanalyse Experten aus dem Bereich Foresight im Rahmen einer Interviewreihe zu verfügbaren Anwendungen, den Potenzialen von Big Data sowie den aktuell noch bestehenden Herausforderungen befragt. Ein Abgleich der Ergebnisse aus der Recherche und der Datenerhebung mit den bereits erarbeiteten Anforderungen verdeutlicht Forschungsbedarfe, die im Anschluss an die kritische Reflexion der technologischen Möglichkeiten dokumentiert wurde.

#### 3.1 Verfügbare Systeme aus Expertensicht

Zur Identifikation und Analyse verfügbarer Assistenzsysteme wurde eine kaskadierende Recherche- und Analysestrategien verfolgt:

Zunächst wurde eine Übersicht über ausgewiesene Foresight-Experten aus wissenschaftlichen Einrichtungen, spezialisierten Beratungsorganisationen

und öffentlichen Einrichtungen im deutschsprachigen Raum zusammengestellt. In leitfadengestützten Telefon-Interviews (Dauer ca. 60 Minuten) teilten Experten (N=20) ihre Kenntnisse und Einschätzungen zu bisherigen Möglichkeiten („Stand der Technik“) und Potenzialen mit, die durch Weiterentwicklung der Systeme entstehen können.

In einer Internetrecherche, ergänzt durch Hinweise der befragten Experten und Nachrecherchen in wissenschaftlicher Fachliteratur konnten Assistenzsysteme identifiziert werden, die BigData-Ansätze im Kontext Foresight nutzen. Deren Analyse umfasste die Auswertung (je nach Zugänglichkeit der entsprechenden Informationsbausteine) schriftlicher Dokumente zum jeweiligen System, Angaben aus Interviews mit dem jeweiligen Anbieter und Nutzern (jeweils qualitative Inhaltsanalyse entlang des Forschungsfragenkatalogs) sowie Eindrücke aus einer Probe- oder Demonstrationsanwendung. Im Fokus der Auswertung stand, inwiefern die Systeme mit ihrem Leistungsspektrum die spezifizierten Anforderungen erfüllen, welche Daten zugrunde liegen, wie Ergebnisse generiert und präsentiert werden bzw. wie das jeweilige System auch im Hinblick auf Mensch-Technik-Interaktion und prozessualer Einbindung gestaltet ist.

Eine übergreifende Auswertung und abschließende zusammenfassende Bewertung zeigt Grenzen und Potenziale auf, auf deren Basis sich aktuelle Anwendungsmöglichkeiten weiter Forschungsbedarfe erschließen lassen.

Die Einsichten der befragten Experten lassen sich folgendermaßen zusammenfassen:

Für Anwendungen mit Bezug zu Foresight werden Unternehmen und Institute, die Assistenzsysteme entwickeln und anbieten, von den Experten als dynamische, wachsende Start-Up-Szene beschrieben. Die benannten Systeme fokussieren meist auf Technology-Foresight und werden für eingegrenzte Aufgaben, wie z. B. zum Technologiemanagement, eingesetzt. Viele und insbesondere große Unternehmen nutzen auf ihre Bedarfe zugeschnittene Systeme. Experten betonen, dass Big Data-Anwendungen voraussetzungsreiche und aufwändige Vorhaben sind. Aus wissenschaftlicher Sicht sei eine Ernüchterung bzgl. des erhofften Nutzens beispielhaft entwickelter Systeme spürbar. Dennoch werden Big Data-Anwendungen als spezieller Bereich der empirischen Sozialforschung verortet, der sich stetig erweitert. Ebenso werden Nutzenpotenziale gesehen, die sich mit zunehmender Reife der Systeme immer besser ausschöpfen lassen. Als Entwicklungsstand wird konstatiert, dass derzeit Systeme eine Unterstützung bei der Exploration eingegrenzter Suchfelder bieten, iterative Explorationsstrategien unterstützen, Muster in Datensätzen erkennen, Daten strukturieren und Vorschläge zu zentralen Informationsbausteinen ausgeben können. Darüber hinaus können Ergebnisse mit ansprechender Visualisierung dem Nutzer einen Überblick zu einem ausgewählten Thema verschaffen und ggf. vertiefende Informationen zu Akteuren, Inhalten und inhaltlich verknüpften Instanzen geben. Je nach Themenfeld sind eng gefasste Prognosen im Sinne von Micro-Trends automatisiert generierbar. Nutzenpotenziale (Abbildung 3) zeichnen sich aus Expertensicht zur Umsetzung effizienterer Suchprozesse

ab sowie zur Erschließung neuer, bisher nicht oder nur schwer zugänglicher Inhalte, der Erweiterung des Foresight-Methodenkanons mit neuen Möglichkeiten zum Erkenntnisgewinn und letztendlich der Möglichkeit zur Erweiterung der inhaltlichen und methodischen Expertise von Foresight-Verantwortlichen.

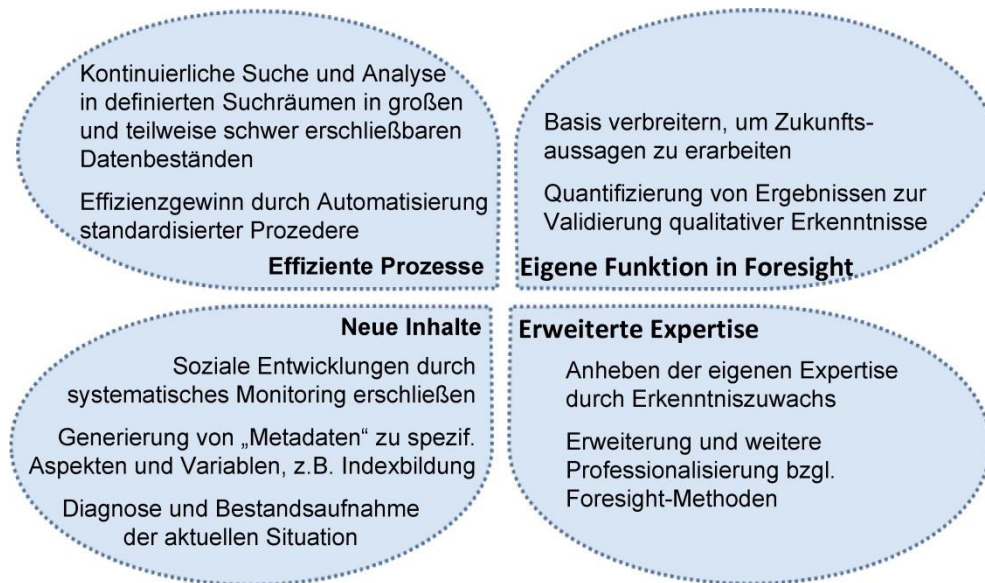


Abbildung 3: Nutzenpotenziale von Big Data für Foresight (eigene Darstellung)

Gleichzeitig sehen die Experten derzeit jedoch auch Grenzen und weitere Entwicklungsbedarfe bei der Anwendung von Big Data-Mining für Foresight, sowohl bei der Gestaltung von Assistenzsystemen und deren Grundlagen, als auch bzgl. der Rahmenbedingungen von Datensammlung, Datenauswertung und Ergebnisverwendung. Herausforderungen und große Aufwände werden in der Auswahl und Erschließung von Datenquellen gesehen, nicht in der prinzipiellen Verfügbarkeit. Als weitere Herausforderung wird ebenso die Konkretisierung und Abbildung der eigenen Fragestellung in einem System gesehen, also das Herunterbrechen einer komplexen Fragestellung in Indikatoren und deren Übersetzung in Algorithmen, bei gleichzeitiger Einhaltung (sozial-)wissenschaftlicher Standards. Zum Beispiel erbringt eine „freie“ Suche nach Themenclustern interpretierbare Ergebnisse bisher nur in theoriegeleiteten, experimentellen Ansätzen in bereinigten und nicht zu großen Datenmengen. Experten weisen darauf hin, es gäbe nach wie vor nicht viele echte Big Data-Anwendungen für Foresight, und betonen die Notwendigkeit menschlicher Fähigkeiten für die (kreativen) Kernaufgaben in einem Foresight-Prozess.

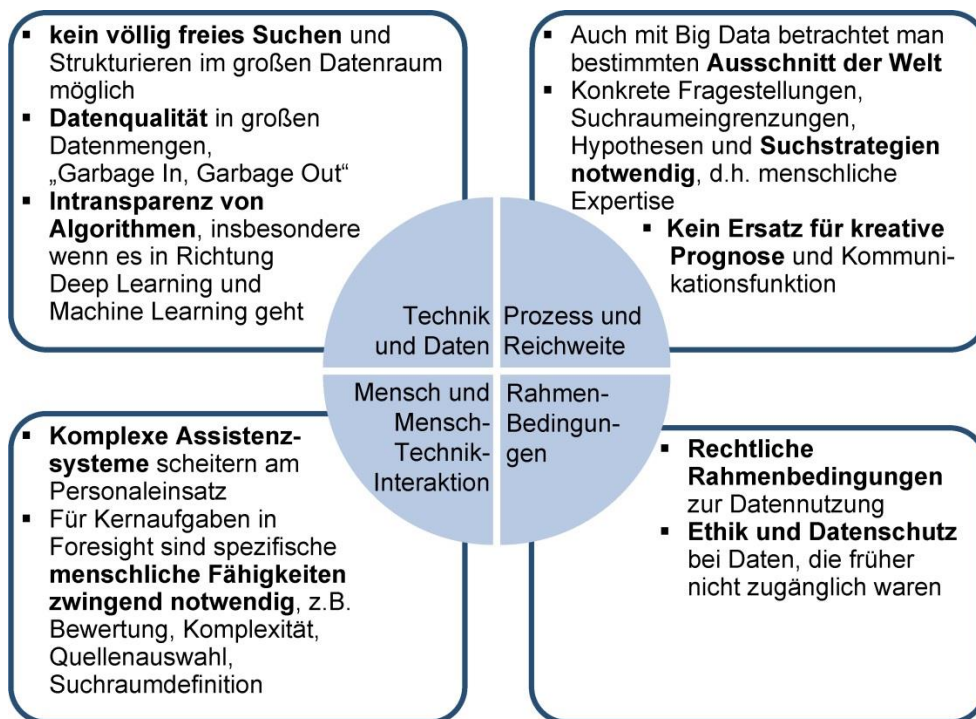


Abbildung 4: Grenzen von Big Data für Foresight (eigene Darstellung)

Ihren Ursprung finden die Big Data-Assistenzsysteme für Foresight in Ansätzen aus unterschiedlichen Kontexten, aus denen aktuell spezifische und hybride Systeme weiterentwickelt werden. Aus der Recherche und Analyse verfügbarer Systeme (ausgewählte Beispiele werden im Text benannt) lassen sich folgende Erkenntnisse zusammenfassen:

Eine Art von Systemen entstammt der Bibliometrie, in der sie zur Auswertung wissenschaftlicher Publikationsdatenbanken eingesetzt werden. Bibliometrische Ansätze werden z. B. in der Technologiefrühaufklärung verwendet. Dies funktioniert gut in Themenfeldern, in denen (noch) überwiegend klassische Publikationsorgane führender Forschungsnationen genutzt werden. Beispielsweise werden dort mit Hilfe von Big Data-Technologien Metriken errechnet, die als Basis für die Zusammenstellung und Auswahl von Forschungsfeldern in Form von Themenübersichten, Trends und Vernetzungen dienen sollen (Bsp.-System: SciVal unter Nutzung von Scopus und ScienceDirect). Weitere Beispiele sind Ansätze von Stelzer et al. (2015), in dem die Verwendung von Daten aus der Bibliometrie in verschiedenen Phasen von Szenario-Prozessen untersucht werden (Bsp.-System: BibMon-Tech), oder von Li et al. (2014), die den Einsatz von Bibliometrie in Zusammenhang mit Technology Roadmapping erforschen. Ein weiteres Beispiel ist der sogenannte „Topic prominence in Science“-Indikator, der über die Dynamik eines Forschungsfeldes Auskunft geben soll (Elsevier 2017).

Eine weitere Art von Systemen, die ursprünglich für die Auswertung spezifischer Daten entwickelt wurde, sind Systeme, die Patentdaten auswerten.

Eine Herausforderung in der Patentometrie ist, weltweit anfallende Patentdaten derart zu harmonisieren und mit weiterführenden Informationen zu verknüpfen, dass über einfache Recherchen in den existierenden Patentfamilien hinaus valide Indikatoren und Indizes gebildet werden können. Ein Beispiel ist der „Patent Asset Index™“ (Bsp.-System: PatentSight) für eine zukunftsbezogene ökonomische Bewertung von Patenten (Ernst, Omland 2011). Darüber hinaus werden Systeme entwickelt, mit denen auch eine textbasierte Suche z. B. nach ähnlichen Patenten in Patentdatenbanken möglich ist (Pötzl 2017) (Weitere Bsp.-Systeme: PatentInspiration, Octimine, PatSeer Pro Edition von Patent iNSIGHT Pro).

Systeme, die verschiedenartige Datenquellen auswerten und die Ergebnisse integrieren, sind in vielen Fällen aus spezifischen Anwendungen heraus entstanden, befinden sich überwiegend in der Erprobungs- oder Start-Up-Phase und werden stetig weiterentwickelt. Dies betrifft sowohl die An- oder Einbindung verschiedenartiger Datenquellen, was mit einem sehr hohen Aufwand an Quellenauswahl, Datenaufbereitung und Datenmanagement bzw. Datenpflege einhergeht, als auch die Entwicklung und Anpassung von Analysetools sowie Visualisierungs- und Reporting-Möglichkeiten. Auswertungslogiken und -algorithmen legen die meisten Anbieter nicht oder nur ansatzweise offen. Die meisten Unternehmen bieten sowohl Service und Beratung auf Basis des eigenen Systems oder eine Implementierung des Systems beim Kunden mit entsprechenden Schulungs- und Support-Angeboten an. Eine Vielzahl von Beratungsunternehmen mit Fokus Technologiemanagement oder explizit Foresight, aber auch andere Unternehmen und Organisationen, betreiben Monitoringsysteme wie z. B. sogenannte „Trendradars“ (Gausemeier, Brink, Buschjost 2010) oder „Trendscouting“ mit eigens entwickelten Systemen und teilweise hohem Personaleinsatz. (Bsp.-Systeme: trommsdorff+düner, Betterplace Lab, ThoughtWorks Technology Radar Vol.16). Insbesondere große Unternehmen und Organisationen haben eigene Systeme entwickelt, um die für sie relevanten Technologiefelder und weltweiten Entwicklungen zu beobachten, Trends zu bestimmen, verdichtete Informationen für Foresight-Prozesse bereitzustellen und daraus Strategien abzuleiten. Ein Beispiel zum Thema Energie findet sich bei Gracht, Salcher, Kerksenbrook (2015) oder bei Lubbaddeh (2016), ein weiteres Beispiel zu einer Vielzahl von Themenfeldern für eine Region bei Buchmann-Alisch, Koglin, Plaue, Walde (2014)

Über reine Trendradars hinaus gehen Systeme, die zusätzliche Schritte eines strategischen Innovationsprozesses adressieren bzw. ganzheitliche Ansätze zur Unterstützung von Foresight-Prozessen verfolgen (Durst, Durst, Kolonko, Neef, Greif 2015), wie z. B. Systeme zur Unterstützung des Technologiemanagements (Bsp.-System: TechnologieRadar des Fraunhofer IAO). Bausteine in diesem Kontext sind u. a. die Nutzung von Netzwerken, in denen Trendscouts weltweit ihre Erkenntnisse zusammenführen und verknüpfen, oder die Auswertung von Innovationsdaten, die dann über interaktive Visualisierungsmöglichkeiten für Kunden auf unterschiedlichen Aggregationsebenen aufbereitet und zur Verfügung gestellt werden (Bsp.-Sys-

teme: Mapegy, Berlin mit Fokus auf globale Innovationsdaten (Publikationen, Patente, Webseiten, Statistiken, Scoutingnetzwerk in 25 Sprachen) bzgl. Technologie- und Wettbewerbsfeldern; QUID Inc., CA, USA mit Fokus auf innovativste Unternehmen). Darüber hinaus gibt es Systeme, die eine Vielzahl der in einem Innovationsprozess verwendeten Methoden, z. B. zur Szenariokonstruktion oder zum Roadmapping, computerunterstützt als Assistenzsystem integrieren und mit den Systemen zur Informationsbereitstellung verknüpfen. Systeme, die aus den Projekten RAHS (D) und KIRAS (A) entstanden sind: ITONICS modular aufgebaut mit Fokus auf Einbindung und Auswertung vieler verschiedener Informationen und Unterstützungssysteme (ins System eingebettete Foresight-Methoden) zu deren Bearbeitung sowie automatisierter Identifikation von Trends und Technologien; 4strat mit Fokus auf Trend-, Innovations-, Risiko- und Szenario-Management in Form eines Cockpits. Die bisher verfügbaren Anwendungen erlauben zum einen das Einbinden vieler verschiedener Informationen sowie die gemeinsame Bearbeitung und Bewertung der Informationen.

### **3.2 Technische Aspekte: eine kritische Reflektion**

Aktuelle technische Unterstützungssysteme für den Einsatz in von Foresight-Prozessen unterscheiden sich hinsichtlich ihrer Reife, aber auch in Bezug auf ihre Generizität bzw. Methodenorientierung. So können beispielsweise generische Suchwerkzeuge genauso Verwendung finden wie Speziallösungen, die umfassendere Prozessschritte auch explizit softwaretechnisch abbilden. Ansatzpunkte für Weiterentwicklungen hinsichtlich der technischen Reife reichen von Forschungsprototypen, die häufig auf einen spezifischen funktionalen Aspekt fokussieren, über Werkzeuge mit einem konsolidierten Funktionsumfang bis hin zu reifen Produkten. Abbildung 5 ordnet einige dieser Techniken hinsichtlich der Dimensionen „Technische Reife“ sowie „Methodische Einpassung“ ein. „Technische Reife“ bezieht sich darauf, wie „ausgereift“ eine (zunächst generische) technische Lösung ist, von der Entwicklung und Erprobung bis hin zur Verfügbarkeit als bewährtes Produkt. „Methodische Einpassung“ gibt Auskunft darüber, inwieweit eine Technik in einem definierten Anwendungskontext erst erprobt oder schon standardisiert und in kommerziellen Produkten benutzt wird.

<b>Einpassung in (Government)-Foresight/ Wissenschaftskommunikation</b>	<i>Standard bei kommerziellen Anbietern</i>		<i>Hochvol. Dokument- und Metadaten-Speicherung und Suche</i>  <i>Themenklassifikation/ Themenwolken</i>	
	<i>Wissenschaftlich- methodisch konsolidiert</i>	<i>Autorennetzwerk- Analyse</i>	<i>Trendanalyse/ Nowcasting</i>  <i>Experten- identifikation</i>	
	<i>Einzelne Studien</i>	<i>Themen- Forecasting</i>  <i>Quellensuche und -Bewertung</i>	<i>Twitter- Analyse</i>  <i>Basistechnologie Mustererkennung in Text und Daten</i>	
		<i>Forschungs- prototypen</i>	<i>Konsolidierte Werkzeuge</i>	<i>Produkte</i>
		<b>Technische Reife</b>		

Abbildung 5: Einschätzung aktueller Techniken hinsichtlich technischer und methodischer Reife (eigene Darstellung)

Zur Speicherung von und Suche in hochvolumigen Dokumentenkorpora, wie sie für den Einsatz in den definierten Anwendungsfälle grundlegend sind, kann auf ausgereifte Technologien sowohl im kommerziellen als auch im Open Source-Bereich zurückgegriffen werden. Sie stehen als eher generische Basistechnologien zur Verfügung, etwa für die Entwicklung eigener Systeme, aber auch eingebettet in spezialisierte Foresight-Anwendungen.

Im Rahmen der Mustersuche bzw. des Text Minings in großen Dokumentenbeständen stehen Klassifikationsverfahren und Algorithmen zur Berechnung von Dokumentähnlichkeiten auf Produktniveau zur Verfügung. Damit können entsprechende Themenlandkarten erstellt werden. Die genaue Ausgestaltung und Parametrisierung dieser Verfahren ist jedoch vom jeweiligen Anwendungsfall abhängig. Um die inhaltlich-textuelle Ähnlichkeit von Dokumenten und Dokumentmengen besser für Endnutzer zu veranschaulichen, kann auf eine Vielzahl von Visualisierungsmetaphern und -techniken zurückgegriffen werden. Auch hier ist es vom jeweiligen Anwendungskontext abhängig, welche Visualisierung nützlich ist und die inhaltliche Struktur angemessen kommuniziert.

Grundlegende Techniken zur Autorennetzwerkanalyse sind gut erforscht und Bestandteil der meisten bibliographisch orientierten Systeme, etwa wissenschaftlichen Zitationsdatenbanken. Es gibt jedoch wenig standardisierte



Module, Schnittstellen und Methoden, solche Analysen leicht in neu zu entwickelnde Rechercheumgebungen oder andere Quellenarten einzubinden. Allgemeinere Methoden der sozialen Netzwerkanalyse werden typischerweise auf Daten aus den Sozialen Medien (etwa Twitter oder Blogging-Plattformen) erforscht. Während hier zwar konsolidierte algorithmische Grundlagen existieren, gibt es kaum allgemein verwendbare Werkzeugbibliotheken, die eine leichte Einpassung in neue (Foresight-)Anwendungen ermöglichen. Auch methodisch gibt es noch keine umfassende Einschätzung der Verwendung und Nützlichkeit für Foresight-Einsatzszenarien.

Bei der Trendanalyse kann zwischen kurzfristiger und eher langfristiger Vorhersage unterschieden werden. Während sich kurzfristig Trends durchaus gut mit einer gewissen Wahrscheinlichkeit aus der Vergangenheit extrapolieren lassen, gibt es für langfristiges Forecasting kaum ausgereifte Techniken und Werkzeuge. Dies gilt vor allem, wenn neue, kreative Aspekte zu betrachten sind, wie es beim Foresight zentral ist. Allerdings können die oben genannten Techniken des Text-Minings (Suche, Klassifikation, etc.) verwendet werden, um in entsprechenden Quellen zukunftsorientierte Aussagen zu finden. In diesem Falle macht das technische System nicht selbst die Extrapolation, sondern deckt entsprechende (menschliche) Aussagen in diesen Quellen auf. Für beide Arten der Unterstützung zum Themen-Forecasting (Extrapolation und Verdichtung zukunftsbezogener Aussagen) gibt es zwar grundlegende Techniken, bisher aber weder konsolidierte Werkzeuge noch fundierte Methodiken.

Eine Thematik, die für Nutzung von Big Data im Kontext der strategischen Vorausschau immer wichtiger wird, ist das Aufspüren von neuen Quellen und – damit einhergehend – die Bewertung von Quellen. Hier gibt es bisher weder umfassende methodische noch technische Unterstützung. Vielmehr geschieht das Aufspüren relevanter Quellen bisher eher auf der Basis individueller Expertise. Eine Qualitätssicherung und ständige Weiterentwicklung der Basis an Informationsquellen ist nur schwer möglich, insbesondere im Hinblick auf das schnelle Wachstum und die hohe Dynamik der elektronisch zur Verfügung stehenden Daten. Wie weiter oben ausgeführt, ist die Datenbasis Kernpunkt für den nutzbringenden Einsatz von Big Data-Technologie. Daher liegt ein großes Potenzial für neue methodische Erkenntnisse und technische Werkzeuge in der systematischen Auswahl und Prüfung von Informationsquellen.

## **4. Forschungsbedarfe**

Aus der Potenzialanalyse ergeben sich neben den Potenzialen durch und Herausforderungen für den Einsatz von Big Data-Technologie grundlegenden Forschungsbedarfe, mit deren Adressierung der Methodenkanon der Sozialwissenschaftlichen Forschung und der Computational Social Sciences erweitert und ggf. zusammengeführt werden kann:

Datensammlung und Qualitätssicherung der Daten stellen insbesondere bei anfallenden Daten aus neuartigen Quellen wie Social Media oder Verhaltensdaten (z. B. aus „Wearables“ oder Überwachungssystemen) im Hinblick auf Kontrolle der Entstehungs- und Erhebungssituation und Manipulationsmöglichkeiten eine Herausforderung dar. Fragen sind zum Beispiel: Wie aussagekräftig sind die Informationen? Was kann man aus ihnen ableiten? Wie exakt bzw. passgenau ist das Abbild der öffentlichen Meinung in den Sozialen Medien? Im Detail sind hier u. a. Ansätze zur Identifikation von Fake News, Bots, Manipulationen an Datensätzen und Erkenntnisse zum Entstehungszusammenhang von Daten wie Selbstselektion, Freiwilligkeit, soziale Erwünschtheit etc. von Interesse. Fehler-Möglichkeiten zu erforschen und Wege bzw. Leitlinien zu einem sensitiven Umgang mit dieser Art von Daten zu beschreiben, kann einen professionellen Umgang mit Big Data bzw. eine angemessene Ausschöpfung der Potenziale von Daten und Informationen aus sozialen Medien und Verhaltensdaten fördern.

Darüber hinaus ergeben sich bei der Datenauswertung weitere Forschungsbedarfe insbesondere bei der Umsetzung von Machine Learning unter den Anforderungen an wissenschaftliche Vorgehensweisen, d. h. theorie- und hypothesengeleitete Herangehensweisen. Hierbei sind bspw. Fragen zu klären, wie Transparenz, Reliabilität und Validität von Prozeduren und Ergebnissen sicherzustellen sind. Zu erforschen wäre in diesem Kontext ebenso, wie Strukturierung und Selektion der Ergebnisse „lernender Systeme“ sinnstiftend umgesetzt und wie Systeme an sich wandelnde Suchfelder und Bedeutungswandel von Begriffen (automatisiert) angepasst werden können. Des Weiteren gilt es die Möglichkeiten zur technischen Umsetzung von „Bedeutungserteilung“ auszuloten, bei denen u. a. Visualisierungsmöglichkeiten großer Informationsmengen von Interesse sind. Sinnvoll erscheinen in diesem Kontext eine Erprobung in unterschiedlichen Anwendungskontexten und die Übertragung wissenschaftlicher Erkenntnisse aus anderen Forschungsfeldern, wie z. B. zu Wahrnehmung. Damit können schrittweise Evidenzen geschaffen werden, die dann auf einer Metaebene Einsatzmöglichkeiten und Grenzen von Big Data aufzeigen.

Vor dem Hintergrund des Diskurses einer sinnvollen Mensch-Technik-Interaktion mit digitalisierten und virtualisierten Systemen, erscheint ebenso die Definition erforderlicher Kompetenzen und die Bereitstellung komplementärer Bildungsangebote notwendig. Damit können Systeme sachgerecht und verantwortungsbewusst eingesetzt werden.

## **5. Zusammenfassung**

Das Ergebnis der Potenzialanalyse ist zusammengefasst zum einen, dass Basistechnologien, wie die Zusammenführung großer, heterogener Informationsquellen, die integrierte Suche in solchen Informationsbeständen und das Aufspüren von statistisch signifikanten Mustern, für die gewünschten An-

wendungen in Foresight existieren sowie in spezifischen Anwendungsfeldern und Kontexten schon in Form von Assistenzsystemen umgesetzt werden. Zum anderen wurde deutlich, dass die verfügbaren Systeme noch nicht voll ausgereift und passend für die in den Einsatzszenarien beschriebenen beispielhaften Anwendungskontexte sind. Darüber hinaus zeigt sich Klärungsbedarf zu ethischen und rechtlichen Rahmenbedingungen sowie Forschungsbedarf zur Umsetzung von Big Data-Technologien in unterschiedlichen Anwendungskontexten. Insofern erscheint die Unterstützung von Foresight-Prozessen mit Big Data-Assistenzsystemen als Ergänzung in Kombination mit etablierten und abgesicherten Methoden, abgestimmt in Bezug Methodenmix und prozessualer Einbindung, empfehlenswert.

Insbesondere Systeme zum Technologiescouting und -management sind bereits so weit entwickelt, dass mit ihnen die Aufgaben in Foresight-Prozessen sinnvoll unterstützt werden können. Ein Ersatz etablierter Foresight-Methoden und -Prozessschritte durch eine automatisierte Ergebniserzeugung erscheint aber (noch) nicht möglich. Speziell in den Phasen zu Beginn eines Foresight-Prozesses, d. h. bei der Formulierung der Fragestellungen, Eingrenzung der Suchfelder sowie der Auswahl und Qualitätssicherung der Datengrundlage, erscheint die Einbindung von Menschen mit entsprechender Expertise unabdingbar. Gleiches gilt in den Prozessschritten, in denen übergeordnete Bewertungen und systemische Betrachtungen, z. B. zur Szenario-bildung, angestellt werden müssen.

Allgemein betrachtet wurde auch deutlich, dass Big Data allein nicht alle „Probleme“ lösen kann, sondern bei falscher Anwendung sogar neue erzeugt. Eine falsche bzw. ungeeignete Quellenauswahl kann schnell zu sog. „Big Data Bullshit“ (Nijhuis 2017) führen. Damit ist gemeint, dass etwa Bestätigungsfehler auftreten können oder die verwendeten Daten einen Bias aufweisen, der zu einer Verzerrung bei der Berücksichtigung und somit ggf. zu Benachteiligung bestimmter Bevölkerungsgruppen oder gesellschaftlicher Bereiche führen kann. Darüber hinaus können unpräzise Assistenzsysteme schnell zu ungenauen Ergebnissen, die wieder überprüft werden müssen und somit eine Informationsflut zweiter Ordnung auslösen, d.h. immer mehr Analyseergebnisse werden erzeugt und müssen interpretiert werden. Das bedeutet auch, dass Big-Dat-gestützte Systeme gegenüber herkömmlichen sozialwissenschaftlichen Methoden mit Stichprobenziehung etc. nur dann Vorteile haben, wenn sie in sich funktionieren sowie nützliche und verlässliche Ergebnisse liefern. Daher muss von vorneherein bedacht werden, dass es abgestimmte Mensch-Maschine-Systeme braucht, die letztendlich eine überschaubare Menge an Auswertungsergebnissen liefern.

Abschließend konnte im Rahmen des Forschungsvorhabens „Potenzialeinschätzung von Big Data Mining als methodischer Zugang für Foresight und Wissenschaftskommunikation“ eine Vielzahl an Forschungsbedarfen identifiziert werden, die jedoch nicht nur den Bereich Foresight, sondern die Weiterentwicklung des sozialwissenschaftlichen Methodenrepertoires und die Gestaltung und Anwendung von Big Data-Technologien betreffen. Wie

die im Projekt ebenfalls durchgeführten Analysen im Bereich Wissenschaftskommunikation zeigten, könnten Synergien bei Entwicklung und Anwendung technischer Lösungen bzgl. der Auswertung tagesaktueller Informationsquellen bzgl. möglicher zukünftiger Entwicklungen genutzt werden. So ist beispielsweise die Qualität und Validität von Informationen keine originäre Herausforderung der strategischen Vorausschau. Vielmehr birgt z.B. die Möglichkeit, Fake-News verlässlich zu identifizieren einen generellen Mehrwert in vielen Bereichen und Disziplinen. Entsprechend sollten zukünftige Forschungsarbeiten multidisziplinär ausgelegt sein, um diese Synergieeffekte möglichst optimal auszuschöpfen.

## Literaturverzeichnis

- Abbott A (2014) The Problem of Excess. In: Sociological Theory, 32, 1, S. 1-26.
- Ansoff H I (1976) Managing Surprise and Discontinuity – Strategic Responses to Weak Signals, Zeitschrift für betriebswirtschaftliche Forschung, 28, S. 129 ff, zitiert nach o. A. (1981) WiSt – Zeitschrift für Studium und Forschung, Strategische Unternehmensführung und das Konzept der „Schwachen Signale“, Kleine Abhandlungen, 1981, Heft 6,
- Bachmann R, Kemper G, Gerzer T (2014) Big Data – Fluch oder Segen? Unternehmen im Spiegel gesellschaftlichen Wandels. MITP Verlag, Heidelberg.
- Buchmann-Alish M, Koglin G, Plaue M, Walde P (2014) Technologieradar Berlin, Technologiestiftung Berlin (Hg.), [https://www.technologiestiftung-berlin.de/fileadmin/daten/media/publikationen/140730\\_Studie\\_Technologieradar\\_2014.pdf](https://www.technologiestiftung-berlin.de/fileadmin/daten/media/publikationen/140730_Studie_Technologieradar_2014.pdf), zuletzt ausgelesen am 15.11.2017
- Buhl H U, Röglinger M, Moser F, Heidemann J (2013) Big Data: Ein (ir-)relevanter Modebegriff für Wissenschaft und Praxis? In: Wirtschaftsinformatik, 55, 2, S. 63-68.
- Danner-Schröder A (2018) Soziale Medien und ihr Veränderungseinfluss auf Organisationen, S. 44. In: Lingnau V, Müller-Seitz G, Roth S (Hrsg.) (2018) Management der digitalen Transformation, Interdisziplinäre theoretische Perspektiven und praktische Ansätze, Vahlen, München, S. 40-62
- Durst C, Durst M, Kolonko T, Neef A, Greif F (2015) A holistic approach to strategic foresight: A foresight support system for the German Federal Armed Forces, technological Forecasting and Social Change 97, 91-104.
- Elkasrawi S, Abdelsamad A, Bukhari S, Dengel A (2016) What you see is what you get? Automatic Image Verification for Online News Content, Proceedings of the International Workshop on Document Analysis Systems, April 11-14, Santorini, Greece, IEEE.

- Elsevier (2017) Now Live! Topic Prominence in Science. URL: <https://www.elsevier.com/solutions/scival/release/topic-prominence-in-science>, zuletzt ausgelesen am 12.10.2017.
- Ernst H., Omland N (2011) The patent Asset Index – A new approach to benchmark patent portfolios, *World Patent Information*, 33(1), S. 34-41.
- Fasel D, Meier A (2016) Was versteht man unter Big Data und NoSQL? In: Fasel D, Meier A (Hrsg) *Big Data: Grundlagen, Systeme und Nutzungspotenziale*, Springer Vieweg, Wiesbaden, S. 3-16.
- Gadatsch A (2017) Big Data – Datenanalyse als Eintrittskarte in die Zukunft. In: Gadatsch A, Landrock H (Hrsg) *Big Data für Entscheider*. Springer Vieweg, Wiesbaden, S. 1-10.
- Gausemeier J, Brink V, Buschjost O (2010) Systematisches Innovationsmanagement. Die Innovations-Datenbank. In: Gundlach C, Glanz A, Gutsche J (Hrsg) *Die frühe Innovationsphase: Methoden und Strategien für die Vorentwicklung*. Symposium, Düsseldorf. S. 471-488.
- Von Gehlen D (2011) Welt ohne Gegenmeinung, wie Google und Co. uns andere Standpunkte vorenthalten, *Süddeutsche Zeitung*, 28. Juni 2011, <http://www.sueddeutsche.de/digital/wie-google-und-co-uns-andere-standpunkte-vorenthalten-welt-ohne-gegenmeinung-1.1112983>, zuletzt ausgelesen am 20.02.2018
- von der Gracht H, Salcher M, Graf Kerssenbrook N (2015) Herausforderung Energie. Der Energieführerschein für die Entscheider von morgen. Redline, München.
- HM Government Horizon Scanning Programme (Hrsg) (2014) *Emerging Technologies: Big Data*. A horizon scanning research paper by the emerging technologies big data community of interest. Verfügbar unter: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/389095/Horizon\\_Scanning\\_-\\_Emerging\\_Technologies\\_Big\\_Data\\_report\\_1.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/389095/Horizon_Scanning_-_Emerging_Technologies_Big_Data_report_1.pdf) (25.10.2017).
- Kayser V., Blind K. (2017) Extending the knowledge base of foresight: The contribution of text mining. *Technological Forecasting & Social Change*, Vol. 116, pp. 208-2015.
- Leonard P M, Vaast E (2017) Social Media and their Affordances for Organizing. *A Review and Agenda for Research*, *Academy of Management Annals*, Vol. 11, pp. 150-188.
- Li Y, Zhou Y, Xue L, Huang L (2014) Integrating bibliometrics and roadmapping methods: A case of dye-sensitized solar cell technology-based industry in China. In: *Technological Forecasting and Social Change*, Vol. 97, pp. 205-222.
- Lingnau V, Müller-Seitz G, Roth S (2018) Eine managementorientierte Perspektive auf die digitale Transformation, S. 4. In: Lingnau V, Müller-Seitz G, Roth S (Hrsg.) (2018) *Management der digitalen Transformation, Interdisziplinäre theoretische Perspektiven und praktische Ansätze*, Vahlen, München, S. 3-13.
- Lubbadeh J (2016) Volle Ladung, *Technology Review* 03/2016, S. 28-32.
- McAfee A., Brynjolfsson E. (2012) Big data: the management revolution. *Harvard business review*. Jg. 90, Vol. 10, S. 61-67.

- Meyer W (2017) Kollateralschäden beim Schrotflintenschießen, Big Brother – Big Data – Big Trouble? Vortrag bei der Jahrestagung der DeGEval – Gesellschaft für Evaluation am 21.09.2017 in Mainz, Session A2: Big Data: Große Datenmengen – Großes Potenzial für die Evaluation? Programm und Abstracts verfügbar unter: <http://www.degeval.de/veranstaltungen/jahrestagungen/mainz-2017/programm> (25.10.2017).
- Müller-Seitz G, Reger G (2010) Networking Beyond the Software Code? An Explorative Examination of the Development of an Open Source Car Project. In: *Technovation*, Vol. 30, pp. 627-634.
- Müller-Seitz G, Beham F, Thielen T (2016) Die digitale Transformation der Wertschöpfung. In: *Controlling & Management Review*, 6. Jg., H. 6, S. 24-31.
- Nijhuis M (2017) How to call B.S. on Big Data: A Practical Guide. Verfügbar unter: <http://www.newyorker.com/tech/elements/how-to-call-bullshit-on-big-data-a-practical-guide> (07.10.2017).
- Pariser E (2012) Filter Bubble. Wie wir im Internet entmündigt werden. Aus dem Amerikanischen von Ursula Held, Hanser, München.
- Pötzl M (2017) Semantische Algorithmen in der Patentanalyse, Vortrag, Tag der gewerblichen Schutzrechte 2017, Informationszentrum Patente, Stuttgart, [https://www.patente-stuttgart.de/downloads/tgs/2017/Poetzl\\_octimine\\_TGS2017.pdf](https://www.patente-stuttgart.de/downloads/tgs/2017/Poetzl_octimine_TGS2017.pdf), zuletzt ausgelesen am 15.11.2017.
- Rohrbeck R, Thom N, Arnold H. (2015) IT tools for foresight: The integrated insight and response system of Deutsche Telekom Innovation Laboratories. *Technological Forecast & Social Change*, Vol. 97, pp. 115-126.
- Schatzmann J, Schäfer R, Eichelbaum F (2013) Foresight 2.0 – Definition, overview & evaluation, *European Journal of Futures Research*, 2013, 1, 15, Springer, DOI 10.1007/s40309-013-0015-4
- Scheiermann A, Zweck A (2014) Big Data für die Sozialforschung. Innovations- und Technikanalyse: Kurzstudie, VDI Technologiezentrum (Hrsg), Schriftenreihe Zukünftige Technologien Consulting Nr. 98, Düsseldorf.
- Simon H A (1971) Designing organizations for an information rich world. In: Greenberger M (Hrsg) *Computers, communications, and the public interest*. Hopkins Press, Baltimore. S. 37-72.
- Stelzer B, Meyer-Brötz F, Schiebel E, Brecht L (2015) Combining Scenario Technique and Bibliometrics für Technology Foresight: The case of Personalized Medicine. In: *Technological Foresight and Social Change*, 98, S. 137-156.
- Wellcome Trust (Hrsg) (2012) Intellectual Property Office: Consultation on Copyright, Verfügbar unter: <https://wellcome.ac.uk/sites/default/files/wtvm054838.pdf> (25.10.2017).
- Zhang Y, Zhang G, Chen H., Porter A.L., Zhu D., Lu J. (2016) Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. *Technological Forecasting & Social Change*, Vol. 105, pp. 179-191)

## Benannte Beispiel-Systeme

- 4strat GmbH, <http://www.4strat.de>, zuletzt ausgelesen am 25.09.2017
- Betterplace Lab, <http://www.betterplace-lab.org>, zuletzt ausgelesen am 22.09.2017
- ITONICS, <https://www.itonics.de>, zuletzt ausgelesen am 25.09.2017
- KIRAS, <http://www.kiras.at>, zuletzt ausgelesen am 25.09.2017
- Mapegy, <http://www.mapegy.com>, zuletzt ausgelesen am 25.09.2017
- Octimine, <https://www.octimine.com>, zuletzt ausgelesen am 22.09.2017
- PatentInsight Pro, <https://www.patentinsightpro.com>, zuletzt ausgelesen am 22.09.2017
- PatentInspiration, <http://www.patentinspiration.com>, zuletzt ausgelesen am 22.09.2017
- PatentSight, <https://patentsight.com>, zuletzt ausgelesen am 22.09.2017
- PatSeer Pro Edition, <https://patseer.com/pro-edition>, zuletzt ausgelesen am 25.09.2017
- QUID Inc., <https://quid.com>, zuletzt ausgelesen am 25.09.2017
- RAHS, <https://www.rahs-bundeswehr.de>, zuletzt ausgelesen am 25.09.2017
- ScienceDirect, <https://www.sciencedirect.com>, zuletzt ausgelesen am 15.09.2017
- SciVal, <https://www.elsevier.com/solutions/scival>, zuletzt ausgelesen am 15.09.2017
- Scopus, <https://www.elsevier.com/solutions/scopus>, zuletzt ausgelesen am 15.09.2017
- TechnologieRadar des Fraunhofer IAO, <https://www.iao.fraunhofer.de/lang-de/leistungen/mensch-und-innovation/technologiemangement.html>
- ThoughtWorks Technology Radar, <https://www.thoughtworks.com>, zuletzt ausgelesen am 22.09.2017
- Trendexplorer, <https://www.trendexplorer.com>, zuletzt ausgelesen am 22.09.2017
- trommsdorff+düner, <https://www.td-berlin.com>, zuletzt ausgelesen am 22.09.2017

**Dipl.-Kfm. techn. Christian K. Bosse:** studierte Betriebswirtschaftslehre mit technischer Qualifikation in Informatik an der TU Kaiserslautern und der Auckland University of Technology (Neuseeland). Seit 2011 arbeitet er am Institut für Technologie und Arbeit im Kompetenzfeld „Kooperation, Innovation und Strategie“ und begleitete u. a. als Mitglied des Evaluationsteams den BMBF-Foresight-Prozess.

*Institut für Technologie und Arbeit, Trippstadter Straße 110, 67663 Kaiserslautern, Tel.: +49(0)631-20583-19, E-Mail: christian.bosse@ita-kl.de*

**Magister rer. nat. Judith Hoffmann:** studierte Psychologie an der Karl-Franzens-Universität Graz (Österreich). Seit 2006 arbeitet sie am Institut für Technologie und Arbeit und leitet dort das Kompetenzfeld „Kooperation, Innovation und Strategie“. Als Mitglied des Evaluationsteams begleitete sie u. a. den BMBF-Foresight-Prozess.

*Institut für Technologie und Arbeit, Trippstadter Straße 110, 67663 Kaiserslautern, Tel.: +49(0)631-20583-15, E-Mail: judith.hoffmann@ita-kl.de*

**Dipl.-Inform. Ludger van Elst:** studierte Informatik, Wirtschaftswissenschaften und Mathematik an der Universität Kaiserslautern. Seit 1999 bearbeitet er am Deutschen Forschungszentrum für Künstliche Intelligenz im Bereich Smarte Daten und Dienste sowohl Forschungs- als auch Industrieprojekte mit den Schwerpunkten Informationsmanagement und Big Data.

*Deutsches Forschungszentrum für Künstliche Intelligenz, Trippstadter Straße 110, 67663 Kaiserslautern, Tel.: +49(0)631-20575-1100, E-Mail: ludger.van\_elst@dfki.de*

