

Big Data Methodik und Vorgehen

Rolf Scheuch, Opitz Consulting GmbH

Dieser Artikel beschränkt sich auf Methoden und Vorgehen für die Projekt-Durchführung und vernachlässigt die Betrachtung spezifischer Methoden und Vorgehensweisen bei der technischen Umsetzung beziehungsweise beim Design von Big Data.

Generell umfassen Big-Data-Vorhaben eine große Spannweite an Projektarten und somit auch an sinnvollen Vorgehensweisen. Nachfolgend werden unter einem Big-Data-Projekt sämtliche projektähnlichen Big-Data-Vorhaben verstanden, die als Projekt oder Teilprojekt auf den im Kapitel „Architektur und Technologie“ erwähnten Big-Data-Technologien beziehungsweise -Architekturen basieren. Explizit nicht betrachtet werden übergreifende Projekte wie etwa eine Big-Data-Initiative, ein Big-Data-Programm, ein Big-Data-Competency-Center oder die Big-Data-Governance.

Die klassischen Vorgehensmodelle und Methoden für die Projektdurchführung [1] unterstützen die Lenkung bei der Entwicklung von IT-basierten Lösungen. Hierbei ist es die Aufgabe des klassischen Projektmanagements (PM), über ein geordnetes Vorgehen das im Vorfeld formulierte Sachziel mit den definierten Qualitätsanforderungen zu erreichen. Dabei spielt das formale Ziel des PM zur Einhaltung eines gesetzten Budget-Rahmens, der zeitlichen Laufzeit wie auch der Qualitätsanforderungen eine entscheidende Rolle.

Seitens der Methodik des PM ist es unerheblich, ob das Projekt in kleinere Einheiten zerlegt wird und man iterativ jeweils auf Veränderungen und Anpassungen reagieren kann oder ob man das Projekt als ein Programm führt [2]. Auch diese Sicht auf das PM und das abgeleitete Vorgehen verändert nicht die grundlegenden Dimensionen: definiertes Sachziel, ein festgelegtes Budget und die Qualitätsansprüche, die im Vorfeld bestehen.

Unter den klassischen Vorgehensmodellen versteht man diejenigen Vorgehensweisen, die schon im Planungsprozess allen Beteiligten eines Projekts konkrete Arbeitsanweisungen zur Verfügung stellen und seitens Sachziel, Budget, Zeitrahmen und Qualitätsanforderungen wohldefiniert sind.

Diese sind unter dem Begriff „Wasserfall“ zusammengefasst.

Der Wasserfall liegt auch dem PM bei vielen BI-Vorhaben zugrunde. Gleichwohl hat sich in der BI-Welt eine alternative Vorgehensweise zur Projekt-Durchführung und -Steuerung erfolgreich etabliert: die Nutzung agiler Vorgehensmodelle [3]. Letztlich ist auch bei den agilen Ansätzen das Sachziel durch die User Story wohldefiniert. Die agile Durchführung soll die explizite Ausprägung der Lösungen besser an den Anforderungen der Bedarfsträger, jedoch unter Einhaltung eines festgelegten Budgets und Zeitrahmens, ausrichten. Hierbei ist der Qualitätsanspruch bei einem agilen Vorgehen ebenfalls fest definiert, wie etwa SCRUM über das Definition-of-Done (DoD: festhalten der Fertigstellungskriterien) des Teams zur Erstellung des Produkts, um die Wünsche des Product Owners an Funktionalität, Qualität, Skalierbarkeit etc. zu erfüllen.

Jedoch zeigt sich ein Mehrwert vieler Big-Data-Ansätze – ähnlich den Data-Mining-Ansätzen – in der Ermittlung beziehungsweise Schärfung eines Sachziels. Dieses explorative Vorgehen entspricht einem Lean-Startup-Ansatz [4]. Hierbei werden, über das im explorativen Vorgehen inhärente validierte Lernen, die Entscheidungen beziehungsweise Modelle anhand von Daten kontinuierlich überprüft und geschärft.

Im Folgenden wird der mögliche Lebenszyklus eines Big-Data-Vorhabens betrachtet und aufgezeigt, dass die Aufgaben und Herausforderungen von Big Data sich im Verlauf des Lebenszyklus verändern. Somit liegt es nahe, auch ein adäquates Vorgehen pro Phase zu wählen und somit im Big-Data-Lebenszyklus ein Mix an Vorgehen und Methoden zu verwenden.

Projekt-Arten

Um die unterschiedlichen Ausprägungen der Projekt-Art auf den Lebenszyklus zu verdeut-

lichen, betrachten wir einige Big-Data-Projekte hinsichtlich ihrer betriebswirtschaftlichen Zielsetzung und der benötigten Implementierungsart. In *Tabelle 1* sind die Schwerpunkte dieser Cluster in Bezug auf den Betrachtungsgegenstand „Lebenszyklus“ herausgestellt.

Die unterschiedlichen Schwerpunkte der technischen Umsetzung bei Big-Data-Projekten bedingen somit ein unterschiedliches Vorgehen bei der Steuerung des Projekts. Die dargestellte Spannweite reicht vom Betrieb der Infrastrukturen für eine Labor-Umgebung oder ein Data Lab über die System-Integration zur BI-Welt oder der Applikationslandschaft bis hin zur Entwicklung eigener Informationsprodukte. Um einen nachhaltigen Nutzen der gewonnenen Erkenntnis einer frühen Phase zu einer Innovation beziehungsweise einem Nutzen für das Unternehmen auszubauen, muss in den folgenden Phasen ein operativer Mehrwert für das Unternehmen entstehen und dies kann in der Regel nicht ohne eine Einbeziehung der Unternehmens-IT erfolgen.

Der Big-Data-Lebenszyklus

Das Application Lifecycle Management (ALM) ist „das Management des Assets „Applikation“ über den gesamten Lebenszyklus von der Idee bis zum End-of-Life, mit dem Zweck, die Anwendungssysteme (hier Big-Data-Lösung) zeitgerecht, verlässlich und anforderungsbezogen zu liefern und gleichzeitig den Wertbeitrag der Applikation kontrolliert und an den Bedürfnissen des Geschäfts ausgerichtet zu gestalten.“ [5]

Bei der Betrachtung des Lebenszyklus wird ein fundamentaler Unterschied eines Big-Data-Vorhabens zu den klassischen Vorgehensweisen transparent: der explorative Charakter der Big-Data-Vorhaben in den frühen Phasen des Lebenszyklus [5]. Das eigentliche Sachziel ist in dieser Phase der Gewinn an einer Erkenntnis selber; das Sachziel

für die weiteren Projektphasen richtet sich an dieser gewonnenen Erkenntnis aus. Ferner gestaltet sich das Sachziel und somit das Produkt in Laufe des Projekts immer detaillierter aus. Diese grundsätzliche stetige Verfeinerung und Anpassung entspricht einem iterativ inkrementellen Vorgehen, wie etwa beim Rational Unified Process (RUP) oder dem Spiralmodell von Boehm. Kritikpunkt an diesen Ansätzen ist der „Wasserfall“-Ansatz innerhalb der einzelnen Iterationen. Beide Vorgehensweisen legen ihren Schwerpunkt auf die kontrollierte Weiterentwicklung bei Groß-Projekten. Big-Data-Projekte zeichnen sich eher durch eine hohe Dynamik und kleine bis mittlere Projektgrößen aus. Gleichwohl lässt sich die grundlegende Idee auf ein Vorgehen bei Big Data übertragen.

Eine weitere Herausforderung bei Big-Data-Projekten liegt in der Veränderung der technologischen Schwerpunkte im Laufe des Lebenszyklus. Um einen nachhaltigen Nutzen der gewonnenen Erkenntnis einer frühen Phase zu einer Innovation beziehungsweise einem Nutzen für das Unternehmen auszubauen, muss in den folgenden Phasen ein operativer Mehrwert für das Unternehmen entstehen. Dies kann in der Regel nicht ohne eine Einbeziehung der Unternehmens-IT erfolgen.

Es lassen sich vier grundlegende Phasen identifizieren, die wiederum aus Iterationen bestehen, jedoch einen klaren Übergang zur nächsten Phase besitzen. Im Folgenden werden die typischen Phasen eines Big-Data-Vorhabens, wie auch in *Abbildung 1* dargestellt, erläutert und die Nutzbarkeit unterschiedlicher Methoden beschrieben.

Ein entscheidender Faktor bei der Auswahl der Methodik ist die Eindeutigkeit und Genauigkeit der Definition des Sachziels. Dies kann in der zeitlichen Abfolge auch zu einem sinnvollen Wechsel bei der Auswahl des geeigneten Vorgehensmodells innerhalb der Phasen führen. Bei der nachfolgenden Erläuterung der Phasen (es wurde bewusst auf die abschließende End-of-Life-Phase verzichtet, um den Schwerpunkt auf die Methodik zu erhalten) soll das einzelne Big-Data-Projekt im Mittelpunkt stehen und nicht eine umfassende Big-Data-Initiative im Unternehmen.

Erste Phase: Vorstudie/Proof-of-Value

Meist beginnen Big-Data-Projekte mit einem Assessment, einer Studie beziehungsweise Analyse mit einem abschließenden Proof-of-

Cluster	Schwerpunkte und Herausforderung
Analyse	Ergebnis: Aussage über eine betriebswirtschaftlich relevante Erkenntnis, oft mit einem strategischen Inhalt. Schwerpunkt: Das Vorhaben besteht in der Regel nur aus einer explorativen Phase und eine operative Implementierung und Betrieb ist nur auf die Labor-Umgebung beschränkt. Herausforderung: Der Nutzen der Erkenntnis liegt erst in nachfolgenden Veränderungen der operativen Tätigkeiten. Aufwand/Ertrag von Big Data ist (wenn überhaupt) nur mit hohem zeitlichen Verzug nachweisbar.
Intelligente/neue Produkte	Ergebnis: Erkenntnis wird Basis eines neuen Geschäftsmodells, das auf einem (Informations-) Produkt für Dritte basiert. Schwerpunkt: Operative Implementierung der Modelle und Technologie durch eine Produktentwicklung. Hierbei muss die Labor- sukzessive in eine Produktiv-Umgebung überführt werden. Herausforderung: Einbindung der IT zur Entwicklung eines neuen, noch unsicheren, aber IT-basierten Geschäftsmodells.
Analytische Quelle	Ergebnis: Erkenntnis dient als weitere Quelle für die bestehende BI-Welt. Schwerpunkt: Produktivsetzung der analytischen Verfahren mittels Big Data und anschließende Integrationsleistung bei Einbindung der Ergebnisdaten in die Datenbewirtschaftung von BI. Herausforderung: Verstärkte Einbindung der IT (BI-Entwickler) bei Entwicklung, Wartung und Betrieb.
Analytisches System	Ergebnis: Erkenntnis wird zu einem eigenen dispositiven (Informations-)System ausgebaut. Schwerpunkt: Aufbau eines Produktivsystems aus der Labor-Umgebung. Herausforderung: Verstärkte Einbindung der IT bei der Entwicklung und im Betrieb.
Effizientere Prozesse	Ergebnis: Erkenntnis wird als Algorithmus in den operativen Systemen verankert. Schwerpunkt: Verankerung der Modelle und Algorithmen von Big Data in den operativen Systemen, wobei dies sowohl klassische Applikationen wie auch Event- beziehungsweise Data-Stream-orientierte Systeme sein können. Herausforderung: Software-Entwicklungsprojekt der IT mit Big-Data-Techniken.

Abbildung 1

Value, um eine vage Idee zu erproben und in der Folge ein Sachziel zu formulieren. Ziel ist es, neben der technischen Machbarkeit vor allem eine unternehmerische Idee und deren wirtschaftlichen Mehrwert zu evaluieren. Sinnvoll erscheint hier ein exploratives Vorgehen beziehungsweise ein klassisches Lean-Startup-Vorgehen, das auch bei den Data-/Text-Mining-Ansätzen Verwendung gefunden hat. Als Ergebnis liefert die Projektgruppe neue Erkenntnisse, die zur Formulierung eines Sachziels (inklusive der Beschreibung eines möglichen wirtschaftli-

chen Nutzens) führen. Im Big-Data-Kontext ist dies meist eine wirtschaftlich relevante Aussage beziehungsweise mathematisch-statistisches Modell.

Anschließend werden die Ergebnisse bewertet und häufig durch einen Proof-of-Value der erwartete Mehrwert dargestellt; im positiven Falle machen die Ergebnisse eine weitere Verfolgung des Sachzieles sinnvoll. Das Motto „fail fast, fail early“ kann auch dazu führen, dass ein Vorhaben mit einem negativen Ergebnis abgebrochen wird oder ein komplett anderes Sachziel als sinnvoll

nachweist. Das Ergebnis dieser Phase ist meist eine Erkenntnis durch die Datenanalyse oder ein Modell zur Simulation von Zusammenhängen.

Zweite Phase: Pilot

Bei einer positiven Bewertung des nun formulierten Sachziels aus der explorativen Phase folgt meist eine Pilot-Phase. Mit einer noch recht vagen Idee versucht man den produktiven Nutzen zu validieren. Als Methodik bieten sich als Vorgehen ein Lean-Startup-Ansatz oder eine agile Methodik an, da beide Ansätze ein bewegliches Sachziel zulassen. Damit der Pilot operativ, wenn auch eingeschränkt, nutzbar ist, wird die IT in die Betriebsorganisation eingebunden. Hierbei durchläuft der Pilot die klassischen Phasen einer Produktentwicklung, jedoch mit einem eingeschränkten Funktionsumfang und dem Ansatz, ein Minimal Viable Product (MVP) [4] zu erstellen.

Als Ergebnis steht eine operativ nutzbare Pilot-Installation zur Verfügung, sie kann seitens der Organisation auf ihre Anwendbarkeit und Vorteilhaftigkeit hin überprüft werden. Mit dem Abschluss dieser Phase sollte die Organisation in der Lage sein, eine Entscheidung über den wirtschaftlichen Nutzen zu treffen und im positiven Fall die weiterführende Produktentwicklung zu beauftragen.

Dritte Phase: Produkt-Entwicklung

In dieser Phase wird man das Sachziel, etwa als User Story oder als Lastenheft, formulieren können und zu einem typischen iterati-

ven, inkrementellen oder agilen Modell bei der Projektdurchführung überwechseln. Die Entscheidung für eine eher agile oder iterative Vorgehensweise hängt von der Komplexität der Integrationsleistung zu operativen Systemen, der Unternehmenskultur selber und der Eindeutigkeit des Sachziels als Basis eines Gewerks ab.

Als Ergebnis der Phase steht nun ein Produkt in einer ersten operativ nutzbaren Qualität zur Verfügung. Erweist sich die Lösung als stabil und werthaltig, kann das Roll-out in der ganzen notwendigen Breite in der Organisation erfolgen. Der Adaption dieser spezifischen Big-Data-Lösung und der Schaffung einer Innovation durch Big Data steht damit nichts mehr im Wege.

Vierte Phase: Betrieb und Wartung

In der Folge werden weitere Anpassungen und neue Ideen durch die Applikationswartung permanent umgesetzt und schnellstmöglich produktiv gestellt. Die Big-Data-Lösung befindet sich nun in der frühen Phase der Applikationswartung mit vielen kleinen Veränderungen. Diese sind nicht unbedingt als neue Release-Stände, die als eigenständige Projekte angesehen werden sollten, zu verstehen, sondern als Requests for Changes (RFC) im Sinne der ITIL. Aus diesem Grunde ist eine Nutzung der agilen Methode des KANBAN bei der Wartung ein guter Ansatz, um die Vielzahl kleinerer RFCs effizient abzuarbeiten.

Wie bei allen Produkten folgt eine ruhigere Phase, die aus Sicht der unternehmerischen Effektivität die Blütezeit der Lösung

darstellt. Die Verantwortlichen der IT betreiben dieses spezifische Big-Data-System in aller Ruhe, der Fachbereich steuert die Lösung im Rahmen eines Produktmanagements und führt gezielt Modernisierungs- und Refactoring-Maßnahmen durch.

Die eigentliche Aufgabe der Organisation ist es nun, ein spezifisches Application Lifecycle Management (ALM) für ihr Big-Data-Vorhaben zu entwickeln, das seitens der Methodik und des Vorgehens den Lebenszyklus und die Zielstellung, wie in der obigen Tabelle ausgeführt, beachtet und nicht dogmatisch an einem Vorgehen über alle Phasen festhält. Insbesondere sind die notwendigen agilen Ansätze zu nutzen, um sukzessive die richtige Big-Data-Lösung zu erstellen und über ein agiles Anwendungsmanagement weiterzuentwickeln [6].

Abbildung 2 zeigt die Veränderung der Mitwirkung unterschiedlicher Organisationseinheiten im Lebenszyklus auf. Diese Kriterien helfen, das eigene Big-Data-Vorhaben einzuschätzen und einen adäquaten Application Lifecycle zu planen.

Data Lab, Governance und Unternehmens-IT

Aktuell fokussieren viele Big-Data-Ansätze allein auf ein Ergebnis, das sich als „Erkenntnisgewinn“ charakterisieren lässt. Die Implementierung von Data Labs ist ein gängiger Ansatz, bei dem Analysten zusammengezogen werden, damit diese Gruppe unabhängig von der Linie-Organisation agieren kann. Das (Big) Data Lab wird als „Forschungsstelle“ angesehen und hat so-

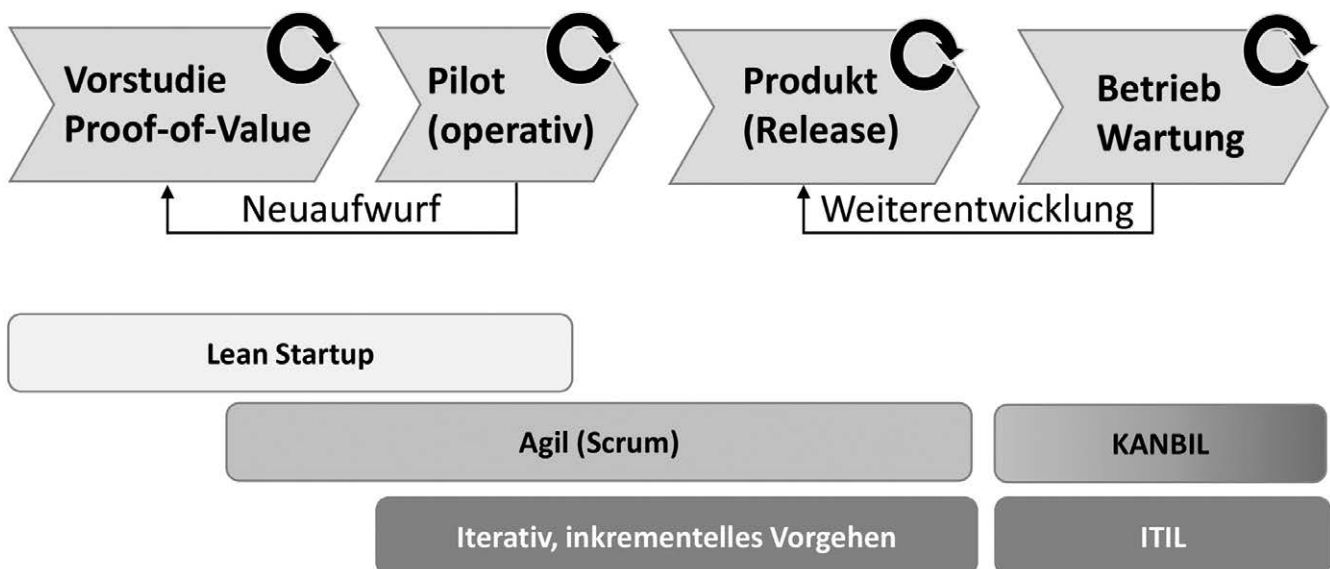


Abbildung 1: Lebenszyklus von Big Data und Vorgehensweisen

mit auch mit ähnlichen Akzeptanzhürden zu kämpfen.

Dieser Ansatz ist vergleichbar mit einem Outsourcing der Forschungsleistung im Unternehmen. Den erkennbaren Vorteilen hinsichtlich gemeinsamer Nutzung einer Big-Data-Infrastruktur und einem Austausch der spezialisierten Analysten steht auch ein grundlegender Nachteil gegenüber.

Mit einem Data-Lab-Ansatz wird es sehr schwierig, den Mehrwert für die Organisation durch Big Data nachzuweisen, da zum einen die Sachziele dieser Gruppe durch die Linienorganisation aufgegriffen werden müssen und zum anderen der wirtschaftliche Nutzen des Data Lab nur indirekt über den Nutzen für die Linienorganisation ermittelt werden kann – und dies oft mit einem erheblichen Zeitversatz.

Die unterschiedlichen Ansätze von Big Data im Unternehmen benötigen eine starke Einbeziehung der Fachbereiche bei Entwicklung und Change Management, um aus Sicht des Fachbereichs die richtige Lösung (oder das richtige Produkt) zu erstellen. Die inhaltliche Führung und Ausrichtung der Big-Data-Forscher, auch in einem Data Lab zentral organisiert, obliegt den Fachberei-

chen, damit eine grundlegende Steuerung der Ausrichtung des explorativen Arbeitens erfolgt. Man verliert eventuell überraschende Ergebnisse eines unabhängigen Forschers, gewinnt aber Stringenz bei den Forschungsinhalten. Ferner steuert das Produkt-/Programm-Management die gesamten Tätigkeiten über alle Phasen des geschil- derten Lebenszyklus – unabhängig von der Form der technischen oder organisatori- schen Implementierung.

Von Bedeutung ist ferner der Abgleich (Governance) mit einem zentralen Competency Center (CC) für Big-Data-Vorhaben im Unternehmen. Zum einen soll das spezifische Big-Data-Vorhaben Rücksicht auf die Big-Data-Strategie nehmen. Zum anderen erscheint es sinnvoll, dass die Big-Data-Vorhaben von den Best Practices, auch bei Auswahl und Betrieb der Big-Data-Infrastrukturen, im Unternehmen profitieren sollen. Die Rolle des CC zur inhaltlichen Steuerung sehen wir in diesem Zusammenhang eher kritisch, da hierbei der explorative und kreative Charakter durch Planungsprozesse behindert und das CC inhaltlich nicht die notwendige fachliche Kompetenz aufbauen wird.

Ferner wird durch das Modell des Big-Data-Lebenszyklus die entscheidende Rolle der Unternehmens-IT deutlich. In den frühen Phasen beschränkt sich die IT auf die Bereitstellung der Infrastruktur und einige Hilfestellungen bei der Daten-Extraktion und im Betrieb. In den späteren Phasen erhöht sich die notwendige Integrationstiefe zur Applikationslandschaft und die IT muss bei der eigentlichen Produktentwicklung eine gewichtigere Rolle einnehmen. *Abbildung 3* beschreibt im Lebenszyklus mögliche Übergabepunkte der technischen Projektverantwortung an die Unternehmens-IT, um das Big-Data-Projekt mit dem nunmehr stark IT-lastigen Schwerpunkt operativ fortzuführen.

Empfehlung

Das Big-Data-Vorgehen entspricht im Wesentlichen dem zyklischen Vorgehen beim Spitalmodell von Boehm, wobei in jedem Zyklus beziehungsweise jeder Iteration das Sachziel beziehungsweise das Ergebnis wieder neu evaluiert und geschärft wird. Jedoch bieten sich auch andere Methoden der Projektsteuerung in den Iterationen an, da Big-Data-Projekte sich von einem explorativen Vorgehen über eine iterative Pro-

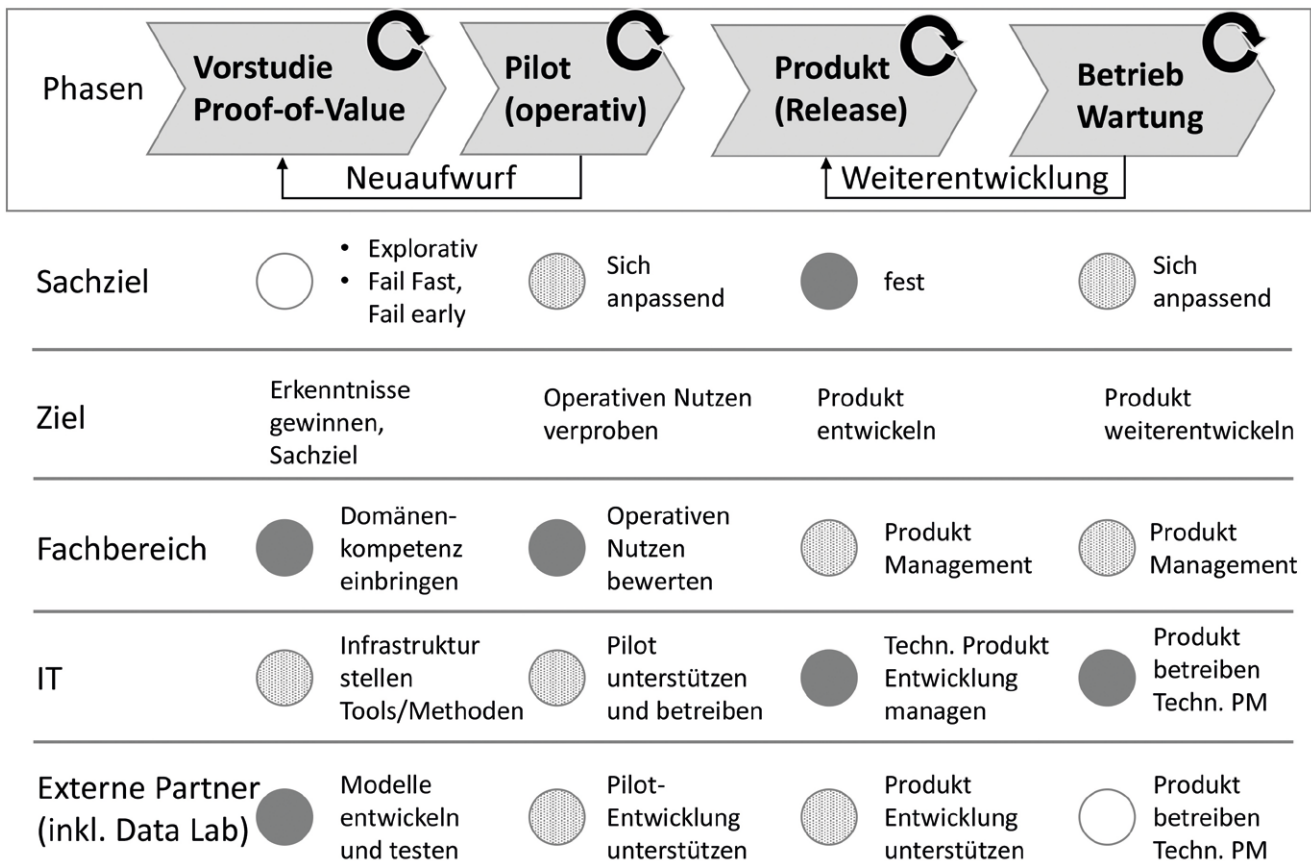


Abbildung 2: Veränderung der Mitwirkung im Lebenszyklus

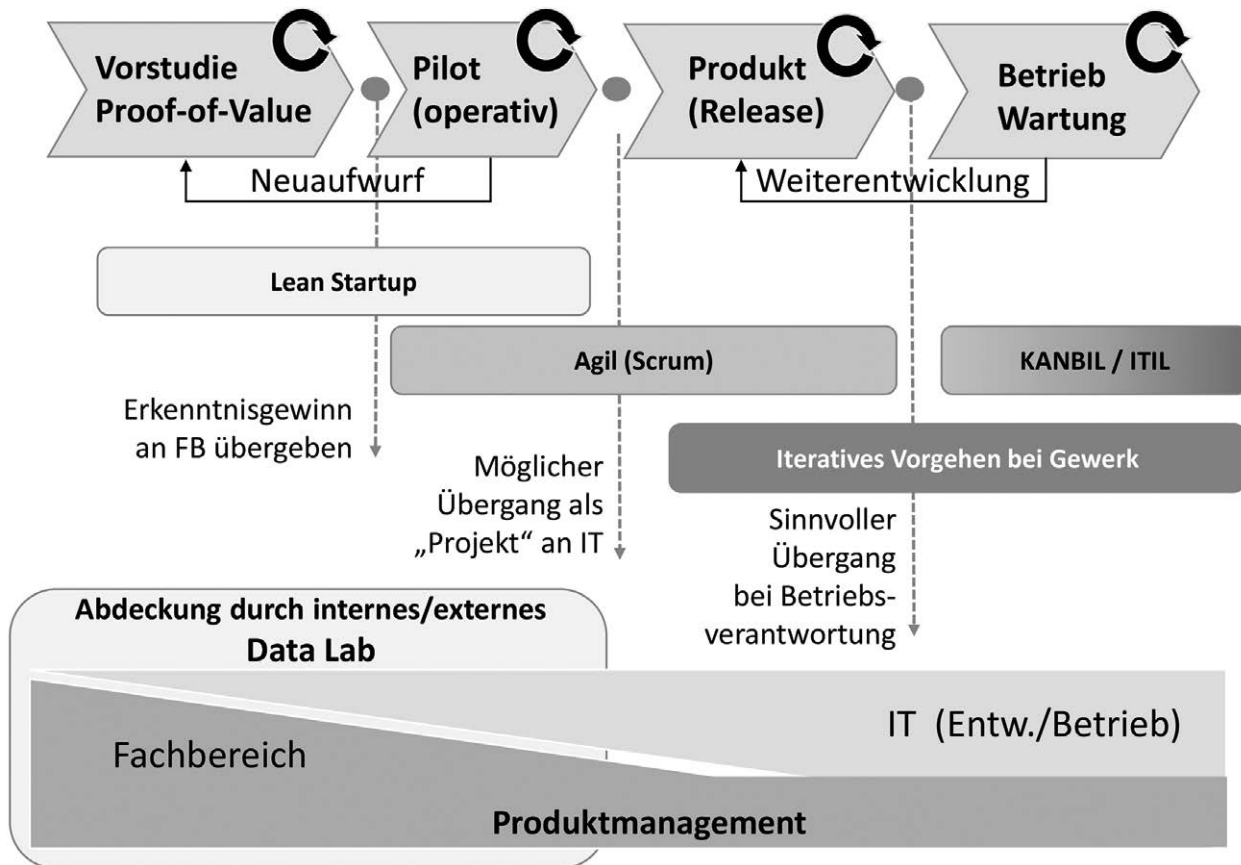


Abbildung 3: Mögliche Übergabepunkte an die Unternehmens-IT

duktentwicklung hin zu einer Betriebs- und Wartungsphase verändern. Jede Phase (mit ihren Iterationen) hat Optionen für die Auswahl einer geeigneten Methodik für die Projektsteuerung.

Einzig bei der frühen Phase eines explorativen Vorgehens erscheint es empfehlenswert, einen Lean-Startup-Ansatz zu wählen. Die Folge-Phasen sollten ein angemessenes Vorgehen und vertraute Methodik verwenden, wobei dies von der Integrationstiefe der Lösung zu den operativen Systemen, der vorherrschenden Unternehmenskultur und der Genauigkeit des Sachziels inklusive der Qualitätsanforderungen abhängt.

Data Labs sind ein effizienter Ansatz, um Big-Data-Kompetenz zu bündeln, jedoch kann dies nur der Startpunkt für die Produktentwicklung und operative Implementierung inklusive des IT-Betriebs sein. Hierbei benötigt man die Kompetenz der Unternehmens-IT, um das (Software-)Produkt zu entwickeln und nachhaltig zu betreiben. Trotz der Bedeutung der Unternehmens-IT bei Entwicklung und Betrieb gehört das Produkt dem Fachbereich und wird durch ein Produkt-Management als

Führungssystem gesteuert. Das Produkt-Management ermöglicht eine Erfolgsmessung des Big-Data-Vorhabens über den gesamten Lebenszyklus.

Die einseitige Ausrichtung der klassischen Methoden und Vorgehensweisen auf Effizienz und Kosten wird dem meist explorativen Charakter der Big-Data-Ansätze nicht gerecht. Effektivität und Veränderung müssen als wesentliche Faktoren im Lebenszyklus Beachtung finden und durch die Verwendungen agiler Methoden und Vorgehen unterstützt werden.

Literatur

- [1] Ferstl, O.-K., Sinz E.: Grundlagen der Wirtschaftsinformatik, 6. Auflage, Oldenburg, 2008.
- [2] PMI, The Standard for Program Management, Pennsylvania (USA), Project Management Institute, Inc., 2006.
- [3] TDWI Germany e.V.: Memorandum für agile Business Intelligence, dpunkt Verlag, 2014.
- [4] Ries, E.: The lean startup: how today's entrepreneurs use continuous innovation to create radically successful businesses. Crown Publishing, 2011, 2014.
- [5] Scheuch, R.: Mehr Effizienz für Big Data – Neue Ansätze für das Applikationsmanagement,

neue Ansätze im Applikationsmanagement, IT-Management September 2015.

- [6] Hüttermann, M.: Agile ALM, Manning, 2012.
- [7] Dittmar, C.: Die nächste Evolutionsstufe von AIS: Big Data. Erweiterung klassischer BI-Architekturen mit neuen Big Data Technologien. In: Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Business-Intelligence-Technologien und -Anwendungen, 5., vollständig überarbeitete Auflage. Unter Mitarbeit von Peter Chamoni: Springer Verlag, S. 55-65.
- [8] Scheuch, R.: Warum versagen typische ALM Ansätze, IM +io Magazin, 2013.

Rolf Scheuch
rolf.scheuch@opitz-consulting.com