

Verbesserte Durchführung von analytischen Projekten mit Hilfe des Data Product Steckbriefes – am Beispiel der CEWE Stiftung & Co. KGaA

Unternehmen, wie die CEWE Stiftung & Co. KGaA, beginnen die ihnen zur Verfügung stehenden Datenmengen zu nutzen, um unternehmerische Mehrwerte zu generieren und dadurch im Wettbewerb bestehen zu können. Beispielsweise werden anonymisierte Daten analysiert, die Informationen zum realen Verhalten innerhalb einer Software beinhalten. Um das Potenzial der Daten wertschöpfend zu nutzen und ein erfolgreiches analytisches Projekt durchzuführen, sind verschiedene Skills unverzichtbar. Die realen Aktivitäten müssen erkannt, interpretiert, bewertet und in den Daten identifiziert werden können. Hierzu bedarf es unternehmerischem Domänenwissen, das oftmals bei den fachlich verantwortlichen Entscheidern liegt. Zudem ist Know-how notwendig, wie die Daten gespeichert, bearbeitet und analysiert werden können, um den Entscheider so zu unterstützen, dass ein unternehmerischer Mehrwert generiert wird. Die benötigten Skills und Verantwortlichkeiten liegen nie bei einer Person, sodass für ein erfolgreiches analytisches Projekt ein Team notwendig ist. Genau an dieser Stelle existieren auch bei der CEWE Stiftung & Co. KGaA Herausforderungen für die erfolgreiche Durchführung eines analytischen Projektes. Die Kommunikation in solchen Projekten ist oftmals unstrukturiert und schwierig zwischen den Beteiligten, da oftmals eine gemeinsame Wissensbasis zwischen Entscheider und Data Scientist fehlt. Zudem ist das weit verbreitete CRISP-DM als Strukturvorlage für Analytik-Projekte zu allgemein. Eine Alternative oder Erweiterung ist notwendig.

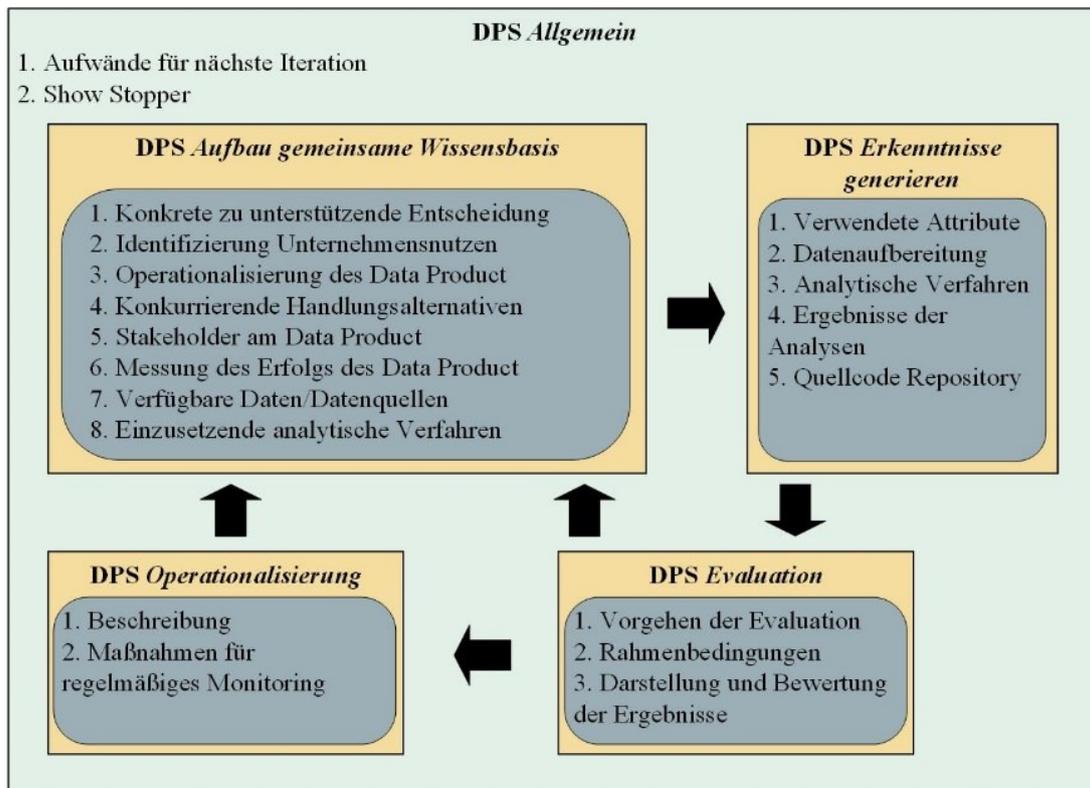


Abbildung 1: Datengetriebener Entscheidungsprozess mit inhaltlichen Auszügen des Data Product Steckbriefes

Kontakt:

Felix Kruse (felix.kruse@uni-oldenburg.de) – Viktor Dmitriyev (viktor.dmitriyev@uni-oldenburg.de)
 Universität Oldenburg – Abt. Very Large Business Application (VLBA)



Um diese Probleme zu lösen, wird das klassische analytische Ergebnis aus dem CRISP-DM als ein Data Product gesehen. Dieses Data Product soll wiederum als Handlungsalternative in einem Entscheidungsprozess berücksichtigt werden. Das Data Product verbindet den Data Scientist, der dieses für die jeweilige Entscheidung entwickelt, und den Entscheider, der ein für die jeweilige Entscheidung optimales Data Product benötigt. An dieser Stelle werden zwei Prozesse miteinander verknüpft. Aus technischer Perspektive, der CRISP-DM, und aus fachlicher Perspektive, der klassische Entscheidungsprozess. Das Ergebnis ist der in Abbildung 1 dargestellte datengetriebene Entscheidungsprozess. Um in diesem Prozess die Probleme der Komplexität, der Kommunikation, der fehlenden gemeinsamen Wissensbasis und der fehlenden Struktur zu lösen, wurde ein Data Product Steckbrief (DPS) entwickelt. Dieser stellt einen Leitfaden für den Durchlauf des datengetriebenen Entscheidungsprozesses dar, um alle relevanten Aspekte zu besprechen und zu dokumentieren.

Der DPS besteht aus fünf Komponenten. Diese werden mit den Beteiligten Data Scientisten und Entscheidern gemeinsam erarbeitet und bieten einen Leitfaden zur Vereinfachung und Strukturierung bei der Entwicklung eines Data Product. Die Komponente *Allgemein* unterstützt bei allen Prozessphasen, um Aufwände und mögliche Show Stopper zu identifizieren und zu dokumentieren. Der Prozess beginnt mit der DPS Komponente *Aufbau gemeinsame Wissensbasis*. In Abbildung 1 sind auszugsweise 8 Ankerpunkte aufgeführt, die in diesem Prozessschritt mit allen Beteiligten definiert und dokumentiert werden sollen. Nach Abschluss des Prozessschrittes folgen die DPS Komponenten *Erkenntnisse generieren*, *Evaluation* und *Operationalisierung* zu denen ebenfalls in Abbildung 1 auszugsweise die relevanten Ankerpunkte, die besprochen und dokumentiert werden sollen, aufgeführt sind.

Der DPS wurde bei der CEWE Stiftung & Co. KGaA in mehreren analytischen Projekten eingesetzt. Dabei behandelten die Projekte inhaltliche Themen in den Bereichen CRM und Produktion, sodass verschiedene fachliche Entscheider eingebunden worden sind. Der Einsatz des DPS führte zu deutlich strukturierteren Besprechungen, indem die Leitfragen des DPS nacheinander besprochen und dokumentiert wurden. Bereits nach dem ersten kompletten Durchlauf des Prozesses mit DPS konnte die Bearbeitungszeit nahezu halbiert werden. Beim erstmaligen Einsatz des DPS mit dem Fachbereich der Produktion konnte der DPS dabei helfen zu erkennen, dass das geplante Projekt mehr als eine Entscheidung betrifft und somit von mehreren Data Products gesprochen wurde. Durch die Reduzierung auf die einzelnen Entscheidungen konnte die Komplexität reduziert und das Projekt deutlich besser strukturiert werden. Bei der Vorstellung des DPS in mehreren Unternehmensbereichen sowie beim Vorstand wurde der Mehrwert der Wissens- und Erkenntnissspeicherung für analytische Projekte hervorgehoben. Ebenso bestätigten die beteiligten Fachbereiche den gemeinsamen Wissensaufbau durch den DPS.