

# Agile Analytics



**Josef Schmid**  
M.A., Inhaber der Dynelytics AG.  
www.dynelytics.com



**Rolf Pfister**  
Dipl. Ing. ETH, Mitglied der Geschäftsleitung der Dynelytics AG.  
www.dynelytics.com

**Die Datenanalyse ist in Unternehmen und Forschungsinstituten in den letzten 20 Jahren immer wichtiger geworden, die Möglichkeiten haben sich enorm entwickelt. Die neuesten Trends sind Agile Analytics und Big Data, die auf Data Lakes zugreifen. Dies sind nicht nur neue Technologien, sondern auch grundsätzlich neue Ansätze für die Herangehensweise bei Datenanalyse-Projekten.**

Agile Analytics ist vor allem ein Ansatz, bei dem die einzelnen Know-how-Träger und deren Teamarbeit entscheidend sind für den Erfolg. Bereits als Unternehmensdaten in relationalen Datenbanken gespeichert werden konnten, war der Informationsgewinn, der durch eine SQL-Abfrage entstand, ein Riesenschritt. Mit zunehmender Komplexität der Datenumgebungen und der Existenz verschiedenster Datenquellen kam es rasch zur Entstehung von Datawarehouses und Datamarts. Diese ermöglichten eine Zusammenführung verschiedener Datenquellen und eine Historisierung der Daten, was die Voraussetzung für genauere und schnellere Analysen ist.

## Die Fallstricke der Analysetätigkeiten

Die wachsenden Analyseaktivitäten zeigten bald auch ihre Grenzen auf: Es kam häufig, je nach Analyseansatz, zu unterschiedlichen oder gar widersprüchlichen Ergebnissen, sozusagen zu verschiedenen Versionen der «Wahrheit». Dies mag zunächst befremdlich erscheinen, ist aber plausibel, wenn man sich die Realität und Komplexität einer Datenhaltung in einem Unternehmen vor Augen hält. Der Schlüssel zu «wahren» Resultaten ist eine einheitliche Definition, und je nach Definition ändern sich auch die Resultate.

Als ein Beispiel sei hier folgendes aufgeführt: Man stelle sich eine Telekommunikationsfirma mit Prepaid-Kunden vor. Eine Account Managerin möchte wissen, wie hoch die Zahl der Kundinnen ist, die zur Konkurrenz gewechselt haben. Dies hört sich zunächst einfach an. Doch wie soll man wissen, dass ein Kunde

zur Konkurrenz gewechselt hat? Es wird ja kein Vertrag gekündigt, keine Umzugsanzeige bekannt gegeben. In einer solchen Situation muss also zunächst der Status «Wechsel» definiert werden. Man kann dies zum Beispiel als Nichtbenutzen des Telefons während dreier Monate festlegen. Wenn aber jemand anderes im Unternehmen plötzlich statt dreier Monate zwei nimmt, tauchen unterschiedliche Kündigerzahlen im Reporting auf.

Um solche Probleme in den Griff zu bekommen, haben sich schnell Methoden und Prozesse wie Master Data Management oder Data Governance etabliert. Damit wurde angestrebt, solche Sachverhalte in einem klaren administrativen Prozess zu definieren und zu implementieren. Dieses Vorgehen hat sich mittlerweile weitgehend durchgesetzt.

## Vordefinitionen erzeugen Schwerfälligkeit

Aber auch hierbei kann es Probleme geben, da solche Prozesse beziehungsweise die entsprechenden Festlegungen zu Inflexibilität und unerwünschter Starrheit führen. Ergänzungen und Änderungen im Datenstamm können dadurch unter Umständen nur sehr langsam realisiert werden. Anpassungen, die aus Businessicht schnell realisiert werden müssten, um auf geänderte Marktverhältnisse zu reagieren, werden dadurch verunmöglicht.

Ähnliche Entwicklungen zeigten sich auch in anderen Bereichen, etwa in der Softwareentwicklung. Grosse Projekte wurden erst bis ins Detail geplant und dann minutiös umgesetzt. Dies hatte



word.cloud © 2015 dynamytics AG

gerade dieser Erkenntnisse ist oft für das Projekt selbst sehr wichtig. Datenaufbereitung sollte daher immer in einem iterativen Prozess geschehen, der auf den lessons learned basiert.

Agile Analytics heisst, dass Anpassungen und Änderungen in sehr kurzer Zeit implementiert und durchgeführt werden können, sei es in einer Business-Intelligence-Umgebung, sei es bei der Entwicklung prädiktiver Modelle. Das Ziel ist, einer Organisation möglichst schnell Informationen und Voraussagen zu liefern, die das sich schnell wandelnde Umfeld einbeziehen.

Um dies zu erreichen, sind organisatorisch, architektonisch und technisch angepasste Rahmenbedingungen nötig. Organisatorisch muss grundsätzlich die Möglichkeit bestehen, Datenzugriffe zu organisieren, die zum Beispiel auch das legislative Umfeld berücksichtigen. Personell müssen entsprechende Ressourcen vorhanden sein, und das technische Umfeld muss ebenfalls up to date sein. Und zu bedenken ist, dass bei Agile Analytics auch Resultate auftauchen können, die leichte Unterschiede aufweisen. Diese müssen dann richtig interpretiert werden, was nur mit entsprechendem Hintergrundwissen möglich ist.

zur Folge, dass manche Projekte massive Kostenüberschreitungen generierten oder sogar scheiterten: Viele Probleme zeigten sich erst später oder während der Entwicklung, wurden aber aufgrund des Planungsansatzes nicht oder nur sehr zurückhaltend berücksichtigt.

Als Konsequenz setzt sich immer häufiger ein flexibles, iteratives Vorgehen durch – man redet hier von agilem Programmieren. Als verbreitete agile Methodologie sei hier SCRUM aufgeführt. Entwicklung ist hier ein Prozess, es geht darum, in umsetzbaren kleinen Schritten zu planen und dann schliesslich aufgrund der sichtbaren Resultate das Vorgehen – im Extremfall täglich – zu revidieren.

### Die Lösung für die Zukunft: Agile Analytics

Ein solches agiles Vorgehen ist auch im analytischen Bereich immer mehr nötig: Ein Vorgehen in kleineren, überschaubaren Schritten, einhergehend mit Überprüfungen und Revisionen des Vorgehens anhand konkreter Resultate. Ein iteratives Vorgehen zwischen Data Scientists und Business Usern wird immer wichtiger. Gefordert wäre also Agile Analytics oder auch Agile Predictive Analytics.

Grundsätzlich kann ein analytisches Projekt in verschiedene Phasen aufgeteilt werden: Definition des Zieles, Datenzugriff, Datenaufbereitung und Modellierung.

Agilität ist aber nicht unbedingt in jeder Phase angebracht. Wenn es zum Beispiel um die Definition eines Zieles geht, ist es empfehlenswert, ein einmal definiertes Ziel beizubehalten. Beispiele für Projekte, die aufgrund veränderter Zielvorgaben schiefgelaufen sind, gibt es genug.

Agilität muss jedoch grossgeschrieben werden in Bezug auf Datenzugriff und Datenaufbereitung, die eigentlich immer die Gewinnung neuer Erkenntnisse zum Ziel haben, und die Nutzung

### Agile Analytics und Big Data verstehen sich dank Data Lakes

Agile Analytics passt sehr gut in die ganze Diskussion um Big Data. Hier wird gemäss Zukunftsvisionen das klassische Warehouse zunehmend durch einen sogenannten Data Lake ersetzt. Damit ist die Kombination verschiedenster Datenquellen «Datensilos» gemeint (interne und externe Daten, einschliesslich Social Media etc.). In einem Data Lake werden Datenobjekte so lange in ihrem ursprünglichen Format gespeichert, bis sie gebraucht werden. Data Lakes sollen Organisationen befähigen, Antworten auf komplexe und dringende Fragestellungen zu finden.

Schon bisher wurde bei vielen Data-Mining-Projekten versucht, durch die Kombination möglichst aller vorhandenen Datenquellen zum Beispiel eine 360-Grad-Kundensicht oder ein «Golden Customer Record» zu erzeugen, die als Grundlage für weitergehende Analysen und Voraussagen dienen.

Der Unterschied zu Agile Analytics ist, dass Agile Analytics und die zugrunde liegende Datenorganisation institutionalisiert sind. In vielen bisherigen Projekten war der Data Lake eine Ad-hoc-Erscheinung. Die Erfahrungen mit Data Mining zeigen aber, dass gerade ein flexibles Vorgehen riesigen Zusatznutzen bringen kann.

Agile Analytics heisst, dass ein Analyseumfeld grundsätzlich flexibel angelegt wird, unter Inkaufnahme von eventuellen Unschärfen. Agile Analytics ist nicht zu verwechseln mit Ad-hoc-Analytics, sondern ist vor allem eine Methodik, die auch auf dem optimalen Zusammenspiel der beteiligten Spezialistinnen und Spezialisten beruht.

Der Nutzen für eine Organisation besteht vor allem in der Schnelligkeit und Flexibilität der Informationsgewinnung und deren adäquaten Umsetzung. 🔍