

Wettbewerbsvorsprung durch Data Maturity

Von Jeremy Peters: GISP, CIMP – Architekt für Stammdatenmanagement- und Datenqualitätslösungen | Leitender Berater | Leitender Ingenieur | Precisely

Zusammenfassung

Data Maturity, auf Deutsch Datenreife, gibt Aufschluss darüber, wo sich Daten befinden, wie sie verwaltet werden, welche Qualität sie haben und welche Art von Fragen sie beantworten. Während Unternehmen die Data-Maturity-Phasen durchlaufen, können sie eine umfassende Analytik-Umgebung realisieren, die eine effektive Analyse von Informationen und damit eine effiziente Entscheidungsfindung in Bezug auf zukünftige Projekte, Märkte und Kunden ermöglicht. Eine umfassende Analytik-Umgebung beinhaltet oft eine integrierte, präzise, konsistente, konsolidierte und angereicherte Übersicht über die Kerndatenbestände innerhalb des gesamten Unternehmens. Diese Datenumgebung kann durch die Automatisierung einer effektiven Datenintegration, Datenbereinigung, Datenanreicherung, Konsolidierung/Objektidentifizierung und des Stammdatenmanagements sowie durch eine deskriptive, prädiktive und präskriptive Datenanalyse geschaffen werden. In diesem Dokument werden sechs Data-Maturity-Phasen beschrieben. Auch wird erläutert, wie Unternehmen in jeder Phase eine effektive Entscheidungsfindung ermöglichen und durch eine zentrale Unternehmenssicht entscheidende Wettbewerbsvorteile erzielen können.

Daten entwickeln sich zu einer bedeutenden Ressource, die aufschlussreiche, vorausblickende Vorhersagen und Empfehlungen ermöglicht. Diese Entwicklung geht mit einer Verbesserung der Rechen- und Speichertechnologie einher, was wiederum leistungsstärkere Analysetools hervorbringt, die von besser qualifizierten Datenmanagement-Profis als Wettbewerbsvorteil genutzt werden. Aber die Menge, Vielfalt und Aktualisierungsrate von Daten steigt schnell. Auch stellt die Qualität oft ein Problem dar, denn die Daten stammen aus unterschiedlichen Quellen: Datenbanken, Webanwendungen, branchenspezifischen Transaktionen und Geräten mit Standortbestimmung, wie Mobiltelefone, sowie unterschiedlichsten Arten von Sensoren. Heute sind Vollzeit-Datenmanagementteams bestehend aus Datenexperten, Analysten und Entwicklern dafür verantwortlich, für ihr Unternehmen eine „Single Source of Truth (SSOT)“, sprich „eine einzige Quelle der Wahrheit“ zu erstellen und zu pflegen. Ihre Aufgabe besteht darin, Datenqualitätsprobleme zu identifizieren und zu beheben, untersuchende, prädiktive und präskriptive Analysen zur Beantwortung dringender Fragen durchzuführen und Analysten in den einzelnen Geschäftsbereichen zu unterstützen.

Das globale Datenwachstum ist durch Vielfalt, Volumen, Geschwindigkeit und durch die dadurch entstehenden Fragen hinsichtlich ihrer Vertrauenswürdigkeit geprägt. Eine Vielzahl neuer Datentypen, die heute zusätzlich zu Stamm- und Transaktionsdaten erfasst werden, führen zu einer neuen Komplexität. Dazu gehören zum Beispiel semistrukturierte (E-Mail, elektronische Formulare usw.), unstrukturierte (Text, Bilder, Videos,

soziale Medien usw.) sowie von Sensoren und Maschinen generierte Daten. Das von Unternehmen und öffentlichen Einrichtungen erfasste Datenvolumen wächst schnell. Diese Datenquellen mit Hunderten von Terabyte oder mehr müssen gehandhabt, gespeichert und analysiert werden. Auch die Frequenz und Geschwindigkeit, mit der Daten – oft in Echtzeit – erstellt, verarbeitet und analysiert werden müssen, steigt in vielen Anwendungen. Die neue Vielfalt, das umfangreiche Volumen und die Frequenz, mit der Daten verarbeitet und analysiert werden müssen, bedeuten noch größere Herausforderungen im Hinblick auf die Vertrauenswürdigkeit und Qualität von Daten. Um Analysen zuverlässig nutzen zu können, müssen Daten bereinigt und ihre Qualität sichergestellt werden. Daten müssen ordnungsgemäß verwaltet, geschützt und genutzt werden, um Marketing und Vertrieb, Customer Experience und Betriebsabläufe besser zu verstehen und zu optimieren.

Data Maturity befasst sich damit, wie Unternehmen ihre Daten verwalten, um wichtige Fragen zu beantworten, die signifikante Auswirkungen auf ihr Geschäftsergebnis haben, zum Beispiel zur Verbesserung der Marketingeffizienz, Reduzierung der Betriebskosten, Umsatzsteigerung, Optimierung der Supply Chain sowie Verbesserung des Produktangebots und des Kundenservice. Anhand der Data-Maturity-Phasen lässt sich die digitale Reife eines Unternehmens in Bezug auf Datenmanagement und Datenanalyse messen. Data Maturity gibt Aufschluss darüber, wo sich Daten befinden, wie sie verwaltet werden, welche Qualität sie haben und welche Art von Fragen damit beantwortet werden.

In diesem Dokument werden die nachstehenden sechs Data-Maturity-Phasen beschrieben. Auch wird erläutert, wie Unternehmen in jeder Phase eine effektive Entscheidungsfindung ermöglichen und durch eine zentrale Unternehmenssicht entscheidende Wettbewerbsvorteile erzielen können.

- Phase 1: Manuelle Datenerfassung und -verwaltung
- Phase 2: Automatische Datenerfassung, -integration und -verwaltung
- Phase 3: Datenqualität
- Phase 4: Stammdatenkonsolidierung
- Phase 5: Zentrale Gesamtansicht des Unternehmens
- Phase 6: Erzielen von Ergebnissen: Datenanalyse zur Verbesserung der Entscheidungsfindung

Phase 1: Manuelle Datenerfassung und -verwaltung

In dieser Phase hat ein Unternehmen noch nicht vollständig die Bedeutung des Managements und der Nutzung von Daten für die Entscheidungsfindung in Bezug auf Funktionen und Anwendungsfälle erfasst. Der Datenbesitz ist auf mehrere Abteilungen aufgeteilt, in denen Kontrolle und Verantwortlichkeit begrenzt sind. Es existiert kein komplettes Inventar des Datenbestands (Data-Dictionary). Mitarbeiter im Bereich Datenmanagement haben begrenzte Kompetenzen und sind für eine begrenzte Anzahl von Aufgaben zuständig, z. B. die Modellierung von Prognosen für den Vertrieb oder das Marketing. Die Erfassung, Verwaltung, Integration und Aufbereitung der Daten für den unternehmensweiten Einsatz erfolgt oft ad-hoc und unkoordiniert anhand unterschiedlicher Systeme, wodurch die Datenqualität erheblich beeinträchtigt wird. Für wichtige Datenbestände gibt es keine Informationssicherheit. Das Reporting ist auf geschäftskritische Aufgaben begrenzt und erfolgt ohne formale Enterprise Information Management (EIM)-Tools oder -Standards, sondern meistens mithilfe von Tabellenkalkulationsprogrammen. Die Tools, die zur Integration und Aktivierung der verschiedenen Daten und Erkenntnisse zum Einsatz kommen, sind nicht optimal. Die Verfahren, mit denen das Unternehmen analytische Methoden auf Daten anwendet, werden gerade erst implementiert. Es bestehen ggf. erste Ansätze für eine datengesteuerte Unternehmenskultur.

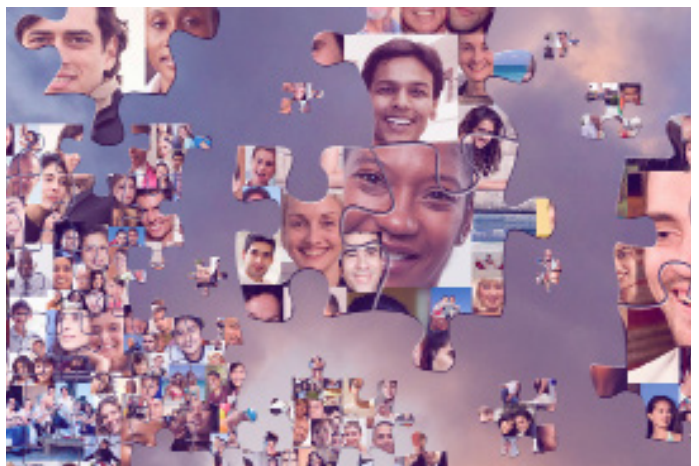
Phase 2: Automatische Datenerfassung, -integration und -verwaltung

In dieser Phase verfügt das Unternehmen über klar definierte, verwaltete und geregelte Data-Warehouse- und Data-Lake-Systeme. Die Planung, Entwicklung und Ausführung von Sicherheitsrichtlinien und -verfahren für die Authentifizierung, Autorisierung, den Zugriff, die Integrität und das Auditing von Daten- und Informationsbeständen wurden eingeführt. In dieser und den folgenden Phasen wird ein leistungsstarkes Data-Governance-Programm mit Kontrollmechanismen und Prozessen eingeführt, um sicherzustellen, dass wichtige Datenbestände unternehmensweit systematisch verwaltet werden.

Außerdem wird ein grundlegendes Enterprise Information Management (EIM)-System als Basis für eine erweiterte analytische Umgebung eingerichtet. Einer der Vorteile von EIM ist seine Fähigkeit, Transformationen bzgl. Datenintegration, Datenqualität und Datenanalysen in grafischen Workflows darzustellen. Diese Transformationen können nun in analytischen Datenbanken und Hadoop umgesetzt werden. Außerdem können Analysen, Regeln, Entscheidungen und Aktionen zu Informationsmanagement-Workflows hinzugefügt werden, um automatisierte Analyseprozesse zu erstellen. Workflows für die gemeinsame analytische Verarbeitung von strukturierten sowie unstrukturierten Daten können eingerichtet und regelmäßig wiederverwendet werden. So können Unternehmen Daten häufiger und schneller verwerten, analysieren und darauf reagieren. Workflows können als Batch-Jobs und Echtzeit-Webservices implementiert werden.

Mit EIM-Data-Discovery-Tools kann ein Unternehmen seine gesamten Datenressourcen erfassen, um ein vollständiges Inventar seiner Datenlandschaft zu erhalten. Es können strukturierte, unstrukturierte und semistrukturierte Daten erfasst werden, um eine den Datenbestand des Unternehmens beschreibende Dokumentationsbibliothek und ein Metadaten-Repository zu erstellen.

Mithilfe von EIM-Datenintegrationstools lassen sich mehrere verschiedenartige Datenquellen integrieren, sprich zusammenführen, filtern und verbinden. Verschiedene Datenquellen, die dieselben Informationen zu einer Entität (z. B. ein Produkt) enthalten, können demselben logischen Modell zugeordnet werden. Damit kann derselbe EIM-Konsolidierungsprozess (Objektidentifizierung) Daten aus verschiedenen Quellen basierend auf einem gemeinsamen Eingabedatenschema verarbeiten. Dieses logische Modell stellt die Business-Objekte und -beziehungen dar, die das Unternehmen für jedes physische Modell bzw. für jede Datenquelle verstehen muss. Diese Datenquellentechnologien können relationale und analytische DBMS (z. B. Oracle bzw. Data Warehouse-Appliances), nicht-relationale Datenmanagementplattformen für Big Data (z. B. Hadoop-Plattform), NoSQL-Datenspeicher (z. B. Graphdatenbanken wie Neo4j), Geschäftsanwendungen (z. B. SAP), Cloud-Anwendungen wie Azure und textbasierte Anwendungen (z. B. XML) enthalten. Diese Datenquellen können über Batch-ETL-Prozesse, Echtzeit-Webservice-Abfragen und/oder Data Federation/Datenvirtualisierung oder andere Integrationsmethoden in ein zentrales logisches Modell integriert werden. Eine automatisierte Datenerfassung, -integration und -verwaltung sorgt für eine bessere Customer Experience und effizientere Betriebsabläufe. Um eine zuverlässige Nutzung von Analysen zu ermöglichen, muss jedoch noch die Datenqualität sichergestellt werden.



Phase 3: Datenqualität

In dieser Phase prüfen Unternehmen die Genauigkeit, Aktualität, Vollständigkeit und Konsistenz ihrer Daten, um ihre Einsatzfähigkeit sicherzustellen. Dabei werden durch automatisches Data-Profiling Datenqualitätsprobleme in Bezug auf Dubletten, Genauigkeit, Aktualität, Vollständigkeit und Konsistenz identifiziert. Anhand der Data-Profiling-Ergebnisse werden Datenqualitätsbewertungen (Data Quality Assessments – DQA) vorbereitet, um Datenqualitätsprobleme, Richtlinien zur Behebung jedes Problems und die Implementierung von Datenqualitätsrichtlinien zu definieren. Danach können die EIM-Datentransformationen so konfiguriert werden, dass die Daten integriert, gepart/normalisiert, standardisiert, validiert und angereichert werden, und zwar basierend auf der Implementierung der im DQA-Bericht beschriebenen Datenqualitätsregeln. EIM-Technologie wird auch zur Standardisierung von Daten bereits bei ihrer Erfassung eingesetzt, wenn z. B. Daten über Onlineformulare in das System eingegeben werden. Beispiel: Bei Eingabe der Adresse in ein Adressformular werden dem Nutzer gültige vordefinierte Adress-Kandidaten zur Auswahl angeboten. EIM-Data-Governance-Technologien und -Prozesse ermöglichen es Dateneignern (Data Stewards), Datensätze, die nicht automatisch oder nicht mit ausreichender Zuverlässigkeit verarbeitet werden konnten, zu überprüfen, zu korrigieren und erneut zu verarbeiten. Anhand der Standardisierung können Datenduplizierung und Dateneingabeprobleme vermieden und Daten unternehmensweit synchronisiert werden.

EIM-Data-Enrichment-Prozesse (Datenanreicherung) steigern den Wert der Daten eines Unternehmens, indem Datensätze um zusätzliche Details ergänzt werden. Daten können z. B. durch raumbezogene Daten, Marketingdaten oder über Daten aus anderen Quellen angereichert werden. Raumbezogene Daten werden oft verwendet, um unterschiedliche Datenquellen anhand eines gemeinsamen Raumbezugs miteinander zu verknüpfen und zusammenzuführen. Beispiel: Adressen können geocodiert werden. Dazu werden die Breiten- und Längenkoordinaten jeder Adresse ermittelt und zusammen mit dem Datensatz gespeichert. Die geografischen Koordinaten können dann für eine Vielzahl von raumbezogenen Berechnungen verwendet werden, z. B. um die dem Kunden nächstgelegene Bankfiliale zu finden. Die gebräuchlichste Methode zur Datenanreicherung besteht darin, gemeinsame Attribute aus einer Datenquelle zusammenzuführen und mit einem Datensatz bestehend aus zusätzlichen Details anzureichern. Eine andere Methode ist ein unscharfer oder exakter Abgleich (fuzzy and exact matching) von ähnlichen Attributen zwischen einer anzureichernden Datenquelle und einem Datensatz mit zusätzlichen Details.

Transliterations- und Übersetzungsfunktionen in EIM-Datenqualitätstechnologien helfen vielen Unternehmen, ihre globale Kommunikation und Information in verschiedenen Sprachen zu standardisieren. Transliterationsfunktionen übertragen Text aus einer Schrift in eine andere, zum Beispiel von der chinesischen Han-Schrift in die lateinische Schrift, ohne jedoch die Wörter selbst zu übersetzen. Mit der Transliterationstechnologie können Eigennamen in nicht-lateinischen Schriften aus verschiedenen Ländern wie China, Taiwan, Hong Kong (Han), Japan (Kanji, Katakana und Hiragana), Korea (Hangul) und Russland (Kyrillisch) im Batchverfahren in die lateinische Schrift umgesetzt werden. Kundenadressen in verschiedenen Sprachen und Schriften können auch mithilfe von globalen Adressvalidierungstechnologien im Batchverfahren gemäß den Standards der Postbehörden transliteriert, analysiert, standardisiert

und validiert werden. Mithilfe von Übersetzungstechnologien lassen sich Informationen, wie Firmennamen im Batchverfahren von nicht-lateinischen Schriften ins Deutsche übersetzen. Diese sprach- und schriftbasierte Standardisierung verbessert nicht nur die Datenqualität, sondern sorgt zudem für eine effektivere Konsolidierung von Entitäten (Person, Ort, Objekt und/oder Gegenstand), um Datensätze zu identifizieren, die dieselbe Entität repräsentieren. Damit entfällt die Verwaltung von unterschiedlichen nichtkonsolidierten Versionen derselben Entität mit unterschiedlichen Informationen in verschiedenen Sprachen und Schriften.

Die Datenbereinigung und -anreicherung verbessert nicht nur die Qualität der Quelldaten. Sie sorgt auch für eine effektivere Stammdatenkonsolidierung (siehe Phase 4), um Datensätze zu identifizieren, die dieselbe Entität repräsentieren. Effektive Datenqualität und Data Governance ermöglichen zuverlässige Analysen, ein besseres Business Monitoring und damit bessere Entscheidungen hinsichtlich zukünftiger Wachstumsstrategien.



Phase 4: Stammdatenkonsolidierung

In dieser Phase erstellen und verwalten Unternehmen eine einheitliche konsolidierte Gesamtansicht ihrer unternehmensweiten Kerndatenbestände mit präzisen und vollständigen Informationen. Dubletten und Unstimmigkeiten werden sowohl innerhalb einzelner als auch zwischen vorhandenen Systemen bereinigt. Die Kerndatenbestände des Unternehmens werden in einer ganzheitlichen konsistenten Ansicht dargestellt und verwaltet. Typische Stammdaten bzw. Stammentitäten betreffen zum Beispiel:

- **Kunden**
- **Produkte**
- **Anbieter**
- **Lieferanten**
- **Unternehmen (Geschäftsbereich)**
- **Mitarbeiter**
- **Geografie (Gebiet/Land)**
- **Standort**
- **Service**
- **Anlagevermögen (z. B. Gebäude)**

EIM-Datenqualitätstechnologien werden zur Automatisierung regulärer Abgleich- und Konsolidierungsprozesse und regelmäßigen Aktualisierungen von Kerndatenbeständen eingesetzt, sobald diese standardisiert und normalisiert sind. Darüber automatisieren Unternehmen Datenabgleichroutinen, um Datenkonsolidierungs- und Duplikationsprobleme zu lösen. In Datenabgleichroutinen können anpassbare Regeln für einen exakten und unscharfen Abgleich verwendet werden. Sie ermöglichen es, eine große Anzahl an effektiven Identifikatoren wie standardisierte Versionen von Produktnamen, -beschreibungen und -kennungen in Produktdatensätzen aus verschiedenen internen Systemen zu vergleichen und zu verknüpfen. Bei diesem erweiterten Fuzzy-Abgleichprozess werden Übereinstimmungen maximiert und die Falsch-positiv-Matches reduziert. Dazu werden für jeden Identifikator basierend auf dessen erwarteter Fähigkeit, eine Übereinstimmung oder Nichtübereinstimmung korrekt zu identifizieren, geeignete Fuzzy-Matching-Algorithmen und Gewichtungen verwendet. Anhand dieser Gewichtungen wird dann die Wahrscheinlichkeit berechnet, dass sich zwei bestimmte Datensätze auf denselben Stammentitätstypen beziehen. Datensätze, die dieselbe Entität darstellen und zuverlässig abgeglichen und miteinander verknüpft sind, lassen sich einfacher zu einem zentralen Stammdatensatz mit angereicherten Informationen konsolidieren. Ein Stammdatensatz kann aus allen übereinstimmenden Entitätsdatensätzen erstellt werden. Dazu werden die Attribute (z. B. der Name) aus jedem Datensatz nach Vollständigkeit, Genauigkeit und/oder Aktualität ausgewertet, um zu ermitteln, welche Attribute aus welchen Datensätzen den Stammdatensatz bilden sollen. EIM-Data-Governance-Technologien und -Prozesse ermöglichen es Dateneignern (Business Stewards), Datensätze, die nicht mit ausreichender Zuverlässigkeit abgeglichen werden konnten, zu überprüfen, zu korrigieren und erneut zu verarbeiten. Danach werden die Stammdaten in geeigneten nachgelagerten Systemen aktualisiert. Die daraus resultierende konsolidierte Ansicht dient als Ausgangspunkt, um Informationen effektiv zu analysieren und Entscheidungen über zukünftige Projekte, Märkte und Kunden zu treffen.

Phase 5: Zentrale Gesamtansicht des Unternehmens

In dieser Phase erstellen und verwalten Unternehmen eine konsistente **zentrale** Gesamtansicht ihrer unternehmensweiten Kerndatenbestände mit präzisen und vollständigen Informationen. Gesamtdarstellungen derselben Datenbestände/ Einheiten werden nicht mehr in verschiedenen Systemen aufbewahrt. Alle Stammdaten werden in ein Stammdaten-Repository geladen und dort verwaltet.

Einige Unternehmen nutzen zum Speichern, Verwalten, Anzeigen, Durchsuchen und Analysieren von Stammdaten und deren komplexen Beziehungen Graphdatenbank-Technologien, um wichtige Zusammenhänge und Trends effektiver zu identifizieren. Graphtechnologien können auch zum Anzeigen, Durchsuchen und Analysieren von Big Data eingesetzt werden, die physisch in einem Nichtgraph-Repository gespeichert sind, zum Beispiel in einer nicht-relationalen Datenmanagement-Plattform wie Hadoop. Auf diese Weise analysieren Banken Milliarden von Transaktionen. Das Speichern der Transaktionsdaten in einem Hadoop-Repository gewährleistet Geschwindigkeit und Skalierung zu niedrigen Kosten, während die Stammkunden- und Kontodaten in einer Graphdatenbank gespeichert und verwaltet werden können. Mittels Integration durch Data Federation/ Virtualisierung, Datenkonsolidierung und Graphdatenbank-Technologien können die in der Graphdatenbank gespeicherten Stammdaten mit den in den ursprünglichen Quellsystemen gespeicherten zugehörigen Informationen verknüpft, visualisiert, durchsucht und analysiert werden. Diese zentrale konsolidierte Gesamtansicht gibt Aufschluss über die Stammdatengruppen eines Unternehmens (z. B. Kunden, Produkte, Anlagevermögen) und über maßgebende und einflussreiche Verbindungen innerhalb und zwischen diesen Gruppen.



Phase 6: Erzielen von Ergebnissen: Datenanalyse zur Verbesserung der Entscheidungsfindung

In dieser Phase erschließen Unternehmen mithilfe von Analytik-Lösungen den Wert großer Datenmengen, um die Entscheidungsfindung zu verbessern. Unternehmen haben entsprechende Prozesse, Managementstrukturen und Technologien, um leistungsstarke Datenmodelle anzuwenden und damit ihre Marketingeffizienz zu erhöhen, Betriebskosten zu senken, Umsätze zu steigern, die Supply Chain zu optimieren und ihr Produkt- und Kundenserviceangebot zu verbessern.

Eine umfassende Analytik-Umgebung, unterstützt durch Technologien wie Künstliche Intelligenz (KI), Machine Learning, Business Intelligence, Location Intelligence und Graphdatenbanken, ermöglicht es Unternehmen, eine ganzheitliche Sicht auf den Markt zu erlangen und zu verstehen, wie sie in diesem Markt wettbewerbsfähig bleiben. Dies umfasst:

- Deskriptive Analytik – ermittelt anhand von Methodologien wie Datenaggregation und Data Mining, was in der Vergangenheit geschehen ist.
- Prädiktive Analytik – ermittelt anhand von Statistikmodellen und Prognosen, was in der Zukunft geschehen wird.
- Präskriptive Analytik – ermittelt anhand von Optimierungs- und Simulationsalgorithmen, wie beste Ergebnisse erzielt und Risiken identifiziert werden können, um bessere Entscheidungen zu treffen.

Künstliche Intelligenz (KI) ermöglicht es Maschinen, aus Erfahrungen zu lernen, sich neuen Eingaben anzupassen und Aufgaben menschenähnlich auszuführen. Machine Learning (maschinelles Lernen), ein Teilgebiet der KI, ist eine Datenanalysemethode, die den Aufbau von Analytikmodellen automatisiert. Machine-Learning-Algorithmen generieren ein mathematisches Modell, das auf Musterdaten (auch Trainingsdaten genannt) basiert. Sie nutzen dieses Modell für Vorhersagen oder Entscheidungen, ohne explizit für die Ausführung dieser Aufgabe programmiert zu sein. Beispiel: Durch Machine Learning aufgebaute Statistikmodelle werden zur Analyse von gewinnrelevanten Kundenverhaltens- und Neigungsanalysen genutzt. Dabei werden alle Verkaufstransaktionen von Kunden analysiert und in Gruppen zusammengefasst, um das Verhalten und die Vorlieben der Kunden vorauszusagen.

Location Intelligence (LI) ermöglicht es, aussagekräftige Erkenntnisse aus raumbezogenen Datenbeziehungen zu gewinnen, um ein spezifisches Problem zu lösen. Raumbezogene Daten werden oft verwendet, um unterschiedliche Datenquellen anhand eines gemeinsamen Raumbezugs miteinander zu verknüpfen und zusammenzuführen. Location-Intelligence-Visualisierung kann helfen, Muster und Tendenzen zu identifizieren, indem Daten mithilfe raumbezogener Analysetools wie etwa thematische Karten und raumbezogene Statistiken in einer Kartenansicht angezeigt und analysiert werden. Location Intelligence ermöglicht es, „Datennadeln in einem Datenhaufen“ zu finden, indem räumliche Beziehungen zum Filtern relevanter Daten genutzt werden. Location-Intelligence-Technologien können Datenverarbeitungs-, Visualisierungs- und Analysetools für die zeit- und raumbezogene Dimension von Daten bereitstellen und dabei helfen, wichtige Erkenntnisse aufzudecken und verwertbare Informationen zu liefern.

Graphdatenbank-Technologien werden zum Speichern, Anzeigen, Durchsuchen und Analysieren von Stammdaten sowie den zugehörigen Entitäten (z. B. Transaktionen) und komplexen Beziehungen eingesetzt, um wichtige Zusammenhänge und Trends zu identifizieren. Viele Unternehmen, mit denen wir täglich interagieren, verwenden Graphdatenbanken und -analysen, wie etwa Google, Facebook, LinkedIn und Amazon. Graphen werden in vielen Formen und Größen und für die unterschiedlichsten Geschäftsaufgaben eingesetzt, zum Beispiel für das Google Website-Ranking, Amazon Market Basket sowie für soziale Netzwerke wie Facebook und LinkedIn. Analysen von sozialen Netzwerken, wie Facebook und LinkedIn, können Erkenntnisse über Personen- oder Unternehmensgruppen sowie über prägende Verbindungen innerhalb und zwischen diesen Gruppen liefern. Technologien, die in der Lage sind, komplexe Beziehungen zu verarbeiten, haben mit dem Big-Data-Wachstum an Bedeutung gewonnen. Richtig konzipiert und ausgeführt, stellen Graphen eine der intuitivsten Möglichkeiten für die Analyse von Informationen dar. Da jedes Subjekt nur einmal mit seinen gesamten Beziehungen im Kontext aller anderen Subjekte und ihren Beziehungen dargestellt wird, können die kompletten Zusammenhänge auf der Ebene des Gesamtbilds festgestellt werden. Anhand von Zentralitätsalgorithmen und Einflussanalysen können die Wichtigkeit und Bedeutung einzelner Entitäten und Beziehungen gemessen werden.

Fazit

Data Maturity gibt Aufschluss darüber, wo sich Daten befinden, wie sie verwaltet werden, welche Qualität sie haben und welche Art von Fragen sie beantworten. Während Unternehmen die Data-Maturity-Phasen durchlaufen, können sie eine umfassende Analytik-Umgebung realisieren, die eine effektive Analyse von Informationen und damit eine effiziente Entscheidungsfindung in Bezug auf zukünftige Projekte, Märkte und Kunden ermöglicht. Eine umfassende Analytik-Umgebung beinhaltet oft eine integrierte, präzise, konsistente, konsolidierte und angereicherte Übersicht über die Kerndatenbestände innerhalb des gesamten Unternehmens. Diese Datenumgebung kann durch die Automatisierung einer effektiven Datenintegration, Datenbereinigung, Datenanreicherung, Konsolidierung/Objektidentifizierung und des Stammdatenmanagements sowie durch eine deskriptive, prädiktive und präskriptive Datenanalyse geschaffen werden. Eine hohe Datenreife ermöglicht es Unternehmen, ihre Marketingeffizienz zu erhöhen, Betriebskosten zu senken, Umsätze zu steigern, ihre Supply Chain zu optimieren und ihr Produkt- und Kundenserviceangebot zu verbessern.

Data-Maturity-Übersicht

	Phase 1: Manuelle Datenerfassung und -verwaltung	Phase 2: Automatische Datenerfassung, -integration und -verwaltung	Phase 3: Datenqualität
Menschen/ Unternehmen Vision und Strategie	<ul style="list-style-type: none"> Begrenztes Bewusstsein im Management, aber kein unternehmensweiter Ansatz Strategische Entscheidungen werden ohne adäquate Informationen getroffen Abhängigkeit von einigen wenigen qualifizierten Personen 	<ul style="list-style-type: none"> Wachsendes Bewusstsein im Management über den Wert von Datenbeständen in bestimmten Geschäftsbereichen Wert von Informationen ist verstanden, aber Strategie ist noch projektorientiert 	<ul style="list-style-type: none"> Wert von Informationen wird in funktionsübergreifenden Projekten erkannt und genutzt
Data Governance	<ul style="list-style-type: none"> Begrenzte Kontrolle und Verantwortlichkeit 	<ul style="list-style-type: none"> Einführung eines ersten Data-Governance-Programms mit Dokumentation der Risiken bzgl. unkontrollierter Datenbestände 	<ul style="list-style-type: none"> Unternehmen setzt Ziele für den Austausch von Informationen im Hinblick auf die Betriebseffizienz Data-Governance-Technologien ermöglichen Dateneignern (Business Stewards) die Überprüfung, Korrektur und erneute Verarbeitung von unzureichend zuverlässig verarbeiteten Datensätzen
Datenarchitektur	<ul style="list-style-type: none"> Keine formale Informationsarchitektur mit Grundlagen, Anforderungen und Modellen, die Teams beim Austausch von Unternehmensinformationen anleiten 	<ul style="list-style-type: none"> Framework für Informationsarchitektur existiert, kann aber nicht um neue Datenquellen oder erweiterte Analytikfunktionen erweitert werden 	<ul style="list-style-type: none"> Vorlagen mit Best Practices für die Informationsarchitektur für Big Data und Analytik wurden definiert und in bestimmten Bereichen angewendet
Metadaten	<ul style="list-style-type: none"> Wenig oder keine Unternehmensmetadaten Keine Richtlinie oder Unternehmensstrategie 	<ul style="list-style-type: none"> IT ist sich der Metadaten bewusst, verwaltet sie aber nicht strategisch Einsatz von EIM-Data-Discovery-Tools, um Datenressourcen des Unternehmens zu erfassen und ein vollständiges Dateninventar zu erhalten 	<ul style="list-style-type: none"> Inkonsistente Datenbestandsidentifizierung Gemeinsame Repositories und Richtlinien auf Abteilungsebene
Daten-integration und -interoperabilität	<ul style="list-style-type: none"> Keine Integrationsstrategie Minimale Datenintegration Jede Integration erfolgt manuell, kundenspezifisch, meistens von Punkt zu Punkt und einmalig 	<ul style="list-style-type: none"> Unternehmen beginnt, Integration als Herausforderung zu betrachten EIM-Datenintegrationstools werden allmählich zur Integration mehrerer unterschiedlicher Datenquellen eingesetzt API- (z. B. REST, SOAP) oder webbasiert 	<ul style="list-style-type: none"> Synchronisierung von APIs Bereitstellung einiger weniger begrenzter Cloud-Konnektoren und -Services Ggf. sind API-Dokumentationen und ein Entwicklerportal verfügbar
Datenqualität	<ul style="list-style-type: none"> Signifikante Datenqualitätsprobleme Reaktivität bzgl. Datenqualität Keine Möglichkeit zur Identifizierung von Datenqualitätserwartungen 	<ul style="list-style-type: none"> Begrenzte Antizipation bestimmter Datenprobleme Erste Datenqualitätserwartungen werden formuliert Einfache Fehler werden identifiziert und gemeldet 	<ul style="list-style-type: none"> Datenqualitätsdimensionen werden identifiziert und dokumentiert Unternehmen beginnt, die Genauigkeit, Aktualität, Vollständigkeit und Konsistenz seiner Daten und damit deren Einsatzfähigkeit sicherzustellen Anhand von Data-Profiling-Ergebnissen werden Datenqualitätsbewertungen (DQA) vorbereitet, um Datenqualitätsprobleme und Richtlinien für deren Behebung zu definieren Verwendung von EIM-Datenqualitäts-Transformationen, um Daten zu integrieren, zu analysieren/normalisieren, zu standardisieren, zu validieren und anzureichern
Referenz- und Stammdaten	<ul style="list-style-type: none"> Vertrauenswürdige Quellen werden manuell verwaltet Silostruktur mit begrenzter Integration 	<ul style="list-style-type: none"> IT-Abteilungen führen abteilungs-übergreifenden Datenaustausch ein, z. B. Stammdatenmanagement 	<ul style="list-style-type: none"> Taktische MDM-Implementierungen mit begrenztem Umfang und für eine bestimmte Abteilung
Datenanalyse/ Messung	<ul style="list-style-type: none"> Reporting beschränkt sich auf geschäftskritische Aufgaben, wird meistens mithilfe von Tabellenkalkulationen realisiert 	<ul style="list-style-type: none"> Das volle Analytik-Potenzial ist noch nicht richtig verstanden, isolierte und Ad-hoc-Aktivitäten, dennoch akzeptable Ergebnisse 	<ul style="list-style-type: none"> Analytik wird genutzt, um Entscheidungsträger über Ereignisse im Unternehmen zu informieren
Datensicherheit	<ul style="list-style-type: none"> Keine Informationssicherheit für wichtige Datenbestände 	<ul style="list-style-type: none"> Planung, Entwicklung und Ausführung von Sicherheitsrichtlinien und -verfahren für Datenbestände sind definiert 	<ul style="list-style-type: none"> Verstärkte Nutzung von Sicherheitstechnologien und Einführung neuer Tools zur Erkennung von Störfällen und für Sicherheitsanalysen

	Phase 4: Stammdatenkonsolidierung	Phase 5: Zentrale Gesamtansicht des Unternehmens	Phase 6: Erzielen von Ergebnissen: Datenanalyse zur Verbesserung der Entscheidungsfindung
Menschen/ Unternehmen Vision und Strategie	<ul style="list-style-type: none"> Management erkennt den Informationsaustausch als entscheidend für die Verbesserung der Unternehmensleistung an und wechselt vom projektorientierten Informationsmanagement zu EIM 	<ul style="list-style-type: none"> Management führt EIM ein, verbreitet und fördert es Unternehmen hat EIM weitgehend eingeführt, insbesondere eine einheitliche Informationsinfrastruktur 	<ul style="list-style-type: none"> Management versteht Daten als Wettbewerbsvorteil und nutzt sie zur Wertschöpfung und Effizienzsteigerung IT-Abteilung versucht, das Informationsmanagement für Benutzer transparent zu machen
Data Governance	<ul style="list-style-type: none"> Governance-Teams, ein formelles Datenqualitätsprogramm und dedizierte Dateneignern (Data Stewards) helfen, Informationen als Unternehmensressource zu verwalten Data-Governance-Technologien ermöglichen Dateneignern (Data Stewards) die Überprüfung, Korrektur und erneute Verarbeitung von unzureichend zuverlässig konsolidierten Datensätzen 	<ul style="list-style-type: none"> Governance-Teams und Lenkungsausschüsse lösen Probleme in Verbindung mit funktionsübergreifenden Informationen 	<ul style="list-style-type: none"> Überwachung und Umsetzung der Informations-Governance ist unternehmensweit automatisiert
Daten- architektur	<ul style="list-style-type: none"> Informationsarchitektur und zugehörige Standards sind klar definiert und erfüllen weitgehend die Anforderungen bzgl. Volumen, Geschwindigkeit, Vielfalt und Vertrauenswürdigkeit 	<ul style="list-style-type: none"> Unternehmens-Informationsarchitektur dient als Leitfaden für das EIM-Programm und stellt einen unternehmensweiten Informationsaustausch zur Unterstützung der Geschäftsstrategie sicher Unternehmen definiert Standards für Informationsmanagement-Technologien 	<ul style="list-style-type: none"> Geschäftsstrategien werden von Informationsarchitektur umfassend unterstützt, um Wettbewerbsvorteile zu schaffen
Metadaten	<ul style="list-style-type: none"> Unternehmen verwaltet Metadaten und beseitigt semantische Unstimmigkeiten, um Wiederverwendung und Transparenz zu fördern 	<ul style="list-style-type: none"> Unternehmen verwaltet Metadaten und beseitigt semantische Unstimmigkeiten, um Wiederverwendung und Transparenz zu fördern 	<ul style="list-style-type: none"> Unternehmen hat Metadatenmanagement und semantischen Abgleich erfolgreich umgesetzt
Daten- integration und -interoperabilität	<ul style="list-style-type: none"> API-Integrationsstrategie wurde implementiert und APIs wurden zugänglich gemacht Ein einzelnes Integrationsstrategie-Team ist für die Bereitstellung von vordefinierten, wiederverwendbaren Workflows verantwortlich Komplexe Integrations-Benutzeroberfläche, Orchestrierungs- und Abstraktionsschicht sind vorhanden 	<ul style="list-style-type: none"> Viele vorkonfigurierte Verbindungen mit Möglichkeiten zur Integration mit Dingen und zur Verteilung von Endpunkten auf führenden Plattformen und Marktplätzen 	<ul style="list-style-type: none"> Unternehmen hat nahtlose Informationsflüsse und Datenintegration über das gesamte IT-Portfolio hinweg erreicht
Datenqualität	<ul style="list-style-type: none"> Unternehmen erstellt und verwaltet eine einheitliche konsolidierte Gesamtansicht seiner unternehmensweiten Kerndatenbestände mit präzisen und vollständigen Informationen Dubletten und Unstimmigkeiten werden regelmäßig sowohl innerhalb einzelner als auch zwischen vorhandenen Systemen bereinigt 	<ul style="list-style-type: none"> Datenqualitätsindikatoren sind definiert An individuelle Leistungsziele gebundene Datenqualitätserwartungen werden berücksichtigt Branchenspezifische Leistungsniveaus werden zur Planung und Definition von Verbesserungszielen verwendet 	<ul style="list-style-type: none"> Framework für Datenqualitäts- und Datenreife-Governance ist im Einsatz und ermöglicht unternehmensweite Leistungsmessungen zur Identifizierung von Möglichkeiten für eine verbesserte systemische Datenqualität
Referenz- und Stammdaten	<ul style="list-style-type: none"> Geschäftslösung bietet eine „Single Source of Truth (SSOT)“ mit Datenqualität nach dem Closed-Loop-Prinzip. Stammdaten werden in geeigneten nachgelagerten Systemen aktualisiert 	<ul style="list-style-type: none"> Stammdatenmanagement bietet eine „Single Source of Truth (SSOT)“ mit Datenqualität nach dem Closed-Loop-Prinzip. Stammdaten werden in ein ZENTRALES Stammdaten-Repository geladen und dort verwaltet Gesamtdarstellungen derselben Datenbestände/Entitäten werden nicht mehr in verschiedenen Systemen aufbewahrt 	<ul style="list-style-type: none"> Unternehmen hat integrierte Stammdatenbereiche und einheitliche Inhalte realisiert
Datenanalyse/ Messung	<ul style="list-style-type: none"> Mittels analytischer Einblicke wird die Wahrscheinlichkeit vorhergesagt, was mit einer bestimmten Geschäftsaktivität geschehen wird 	<ul style="list-style-type: none"> Prädiktive Analytik unterstützt Unternehmen bei der Optimierung der Entscheidungsfindung, um bestmögliche Maßnahmen zur Maximierung des Geschäftswerts zu ergreifen 	<ul style="list-style-type: none"> Deskriptive, prädiktive und präskriptive analytische Einblicke optimieren Geschäftsprozesse und werden, wo möglich, automatisiert Eine umfassende Analytik-Umgebung, unterstützt durch Technologien wie Künstliche Intelligenz (KI), Machine Learning, Business Intelligence, Location Intelligence und Graphdatenbanken, ermöglicht es dem Unternehmen, eine ganzheitliche Sicht auf den Markt zu erlangen und zu verstehen, wie es in diesem Markt wettbewerbsfähig bleibt.
Datensicherheit	<ul style="list-style-type: none"> Sicherheitsrichtlinien und -verfahren sind zusammen mit angemessenen Umsetzungsmechanismen zur Förderung des Bewusstseins und der Compliance im Einsatz 	<ul style="list-style-type: none"> Zentrale Verwaltung aller sicherheitsbezogenen Fragen und Richtlinien 	<ul style="list-style-type: none"> Auf Unternehmenssicherheit abgestimmte Technologiearchitektur Fokus auf Störfallprävention, -erkennung und -bearbeitung Identitätsmanagement und Datensicherheit im Rahmen der Cloud- und Mobile-Computing-Sicherheit



Über Precisely

Precisely ist globaler Marktführer für Datenintegrität. Das Unternehmen verhilft 12.000 Kunden in mehr als 100 Ländern – 90 Prozent davon sind Fortune 100-Unternehmen – zu genauen und konsistenten Daten. Die Precisely-Lösungen für Datenintegration, Datenqualität, Location Intelligence und Datenanreicherung ermöglichen bessere Geschäftsentscheidungen für bessere Geschäftsergebnisse.

Weitere Informationen auf precisely.com