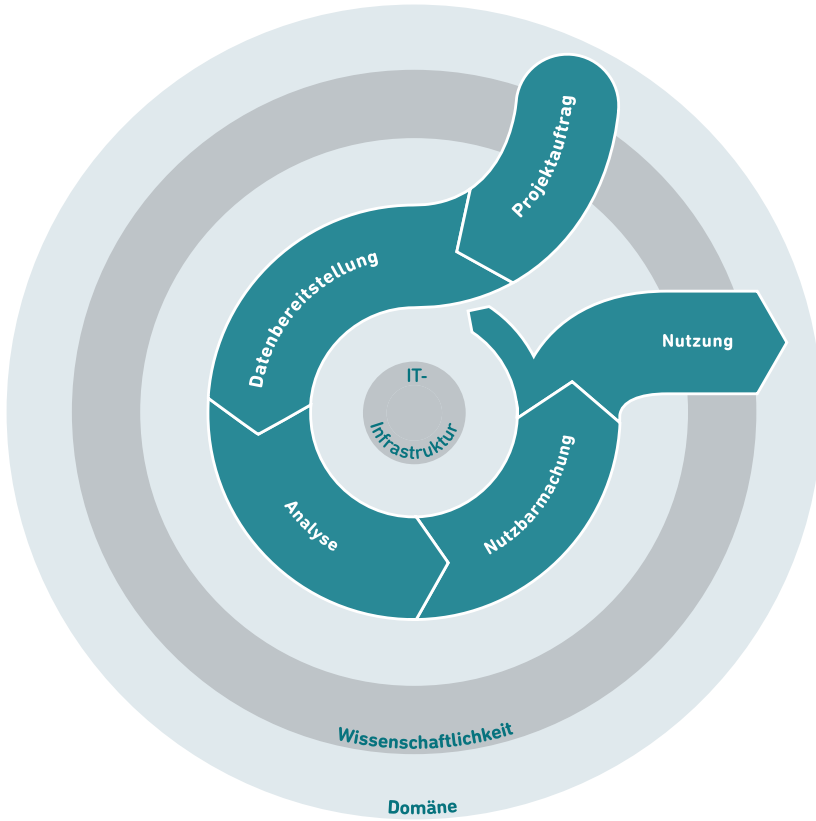


dasc°pm v1.1

Das Data Science Process Modell (DASC-PM) ist ein Vorgehensmodell für Data-Science-Projekte. Es beschreibt die projektrelevanten Schlüsselbereiche und zu durchlaufenden Phasen, erläutert die typischen Aufgaben innerhalb der Phasen und stellt die beteiligten Projektrollen und benötigten Kompetenzen dar.



Domäne
Ein breites Hintergrundwissen in der Domäne ist an vielen Stellen des Data-Science-Prozesses relevant, z. B. bei der Identifikation eines lohnenden Analyseziels, dem korrekten Verständnis von Daten, ihrer Herkunft, Qualität und Zusammenhänge, der Bewertung und Einordnung der erzielten Analyseergebnisse im Kontext der Anwendung sowie der späteren praktischen Nutzung der Ergebnisse. Auch die Beurteilung von Stärken und Schwächen bestehender Lösungen, die fachliche Anforderungsanalyse, die Unterstützung bei der Modellparametrisierung und die abschließende Evaluation des Projekterfolgs werden diesem Bereich zugeordnet. Schließlich lassen sich auch die rechtlichen, gesellschaftlichen und ethischen Aspekte des Data-Science-Projekts an dieser Stelle aufnehmen.

Wissenschaftlichkeit
Wissenschaftlichkeit formuliert keinen Anspruch auf ein vollständig formalisiertes, akademisches, forschungsorientiertes Vorgehen. Während dies im Kontext von Forschungsprojekten so sein kann, bezieht sich der Aspekt der Wissenschaftlichkeit im Business-Kontext vor allem auf eine saubere Methodik, wie sie eben typischerweise als Eigenschaft wissenschaftlichen Arbeitens erwartet wird. Der definierte Projektauftrag ist in jeder einzelnen Projektphase entsprechend methodisch zu bearbeiten. Hervorzuheben sind hier vor allem das Projektmanagement und eine strukturierte Bearbeitung, die bereits durch die Verwendung eines Vorgehensmodells in den Vordergrund gestellt wird. Details zum nötigen Grad an Wissenschaftlichkeit sind unter Berücksichtigung der Projektgegebenheiten und der Domänenspezifika festzulegen.

IT-Infrastruktur
Sämtliche Schritte, die ein Data-Science-Projekt durchlaufen muss, sind von der zu Grunde liegenden IT-Infrastruktur abhängig, das tatsächliche Ausmaß ist allerdings projektindividuell zu bewerten. Auch wenn die Nutzung spezifischer Hard- und Software häufig bereits organisationsintern festgelegt ist, sollten, wenn auch nicht die Auswahl, so doch zumindest die limitierenden und befähigenden Merkmale der IT-Infrastruktur in sämtlichen Projektphasen berücksichtigt werden.

Rollen

Kernrolle „Data Scientist“
Data Scientists sind Spezialisten für den Analysebereich eines Data-Science-Projekts.

Kernrolle „Data Engineer“
Data Engineers kümmern sich um die Beschaffung, Speicherung, Aufbereitung, Strukturierung und Weitergabe von Daten.

Kernrolle „Domänenexperte“
Domänenexperten sind Fachanwender oder Vertreter der Fachanwender.

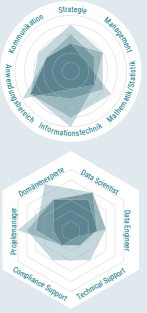
Kernrolle „Projektmanager“
Projektmanager planen, steuern und koordinieren den Gesamttablauf eines Data-Science-Projekts.

Ergänzende Rolle „Technischer Support“
Der technische Support umfasst alle Aufgaben, die erledigt werden müssen, um die technischen Voraussetzungen für die Durchführung des Data-Science-Projekts zu schaffen.

Ergänzende Rolle „Compliance-Support“
Der Compliance-Support ist für die Einhaltung gesetzlicher Vorgaben, die Kompatibilität mit organisationsinternen Regelwerken und das korrekte Verhalten der Projektmitarbeiter verantwortlich.

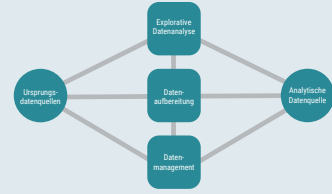
Projektauftrag

Innerhalb einer Domäne bestehende Probleme lösen eine Use-Case-Entwicklung aus. Die vielversprechendsten Use Cases werden anschließend zu einer Data-Science-Projektskizze ausgearbeitet. Alle zugehörigen Aufgaben finden sich in der Phase des Projektauftrags. Durch die frühe, relativ umfassende Betrachtung des Projekts sind hier häufig auch umfassende Fähigkeiten in fast allen Kompetenzbereichen erforderlich.



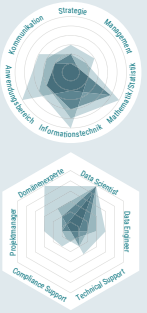
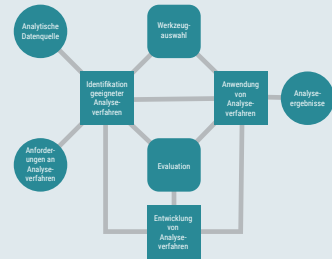
Datenbereitstellung

Innerhalb der Phase der Datenbereitstellung werden die Aktivitäten zusammengefasst, die dem Schlüsselbereich Daten zuzuordnen sind, weshalb der verwendete Begriff weit gefasst ist. Die Phase beinhaltet die Datenaufbereitung (von der Erfassung bis zur Speicherung), das Datenmanagement und eine explorative Analyse. Als Ergebnis dieser Phase entsteht eine für die weitere Analyse geeignete Datenquelle.



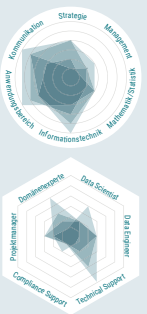
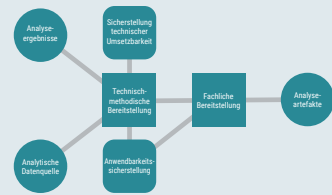
Analyse

In einem Data-Science-Projekt können entweder bestehende Verfahren angewendet oder es müssen zunächst neue Verfahren entwickelt werden – die entsprechende Entscheidung ist eine eigene Herausforderung. Die Phase umfasst daher nicht nur die Analysedurchführung, sondern auch angrenzende Tätigkeiten. Das Artefakt der Phase ist ein Analyseergebnis, das eine methodische und fachliche Evaluation durchlaufen hat.



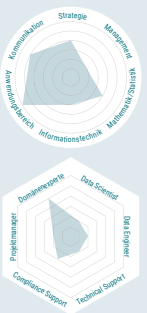
Nutzbarmachung

In der Phase der Nutzbarmachung wird eine anwendbare Form der Analyseergebnisse geschaffen. Projektspezifisch kann dies eine umfangreiche Betrachtung technischer, methodischer und fachlicher Aufgaben bedeuten oder pragmatisch gehandhabt werden. Die Analyseartefakte können sowohl Resultate als auch Modelle oder Verfahren selbst umfassen und werden den Adressaten in unterschiedlicher Form zur Verfügung gestellt.



Nutzung

Die sich an der Projektdurchführung anschließende Verwendung von Analyseartefakten ist nicht als primärer Teil eines Data-Science-Projekts anzusehen. Ein Monitoring ist aber abhängig von der Form der Nutzbarmachung nötig, um die fortbestehende Eignung des Modells in der Anwendung zu prüfen und ggf. Erkenntnisse aus der Nutzung für die Weiter- und Neuentwicklung (auch im Sinne iterativer Vorgehensweisen) zu erlangen.



- Merkmalstragender Bereich
- Kernaufgabe
- Begleitende Aufgabe
- Schnittstellenaufgabe

○ Kompetenzen ○ Rollen



Das Werk ist lizenziert unter einer Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International Lizenz. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Autoren:
Schaik, M., Neuhäus, U., Kaufmann, J., Kälsch, S., Alkazzi, E.M., Bahr, H., Hossain, S., Theunert, R.,
Schaik, U., Bögel, G., Lempertz, C., Jansen, C., Grottel, M., Hübner, S., Hübner, M., von Steiner, P., Fiedler, J.,
Buckmann, J., Schmidt, P., Seyfarth, T., Bredow, W., Wollas, H., Sackmann, S., Göller, P., Weller, F.,
Böhl, C., Siedemann, J., Hübner, U. (2022). DASC-PM V1.1 – Ein Vorgehensmodell für Data-Science-Projekte.
Gefördert durch die NORDAKADEMIE Stiftung.
Quelle: Grafik Wild



Gefördert durch die NORDAKADEMIE Stiftung