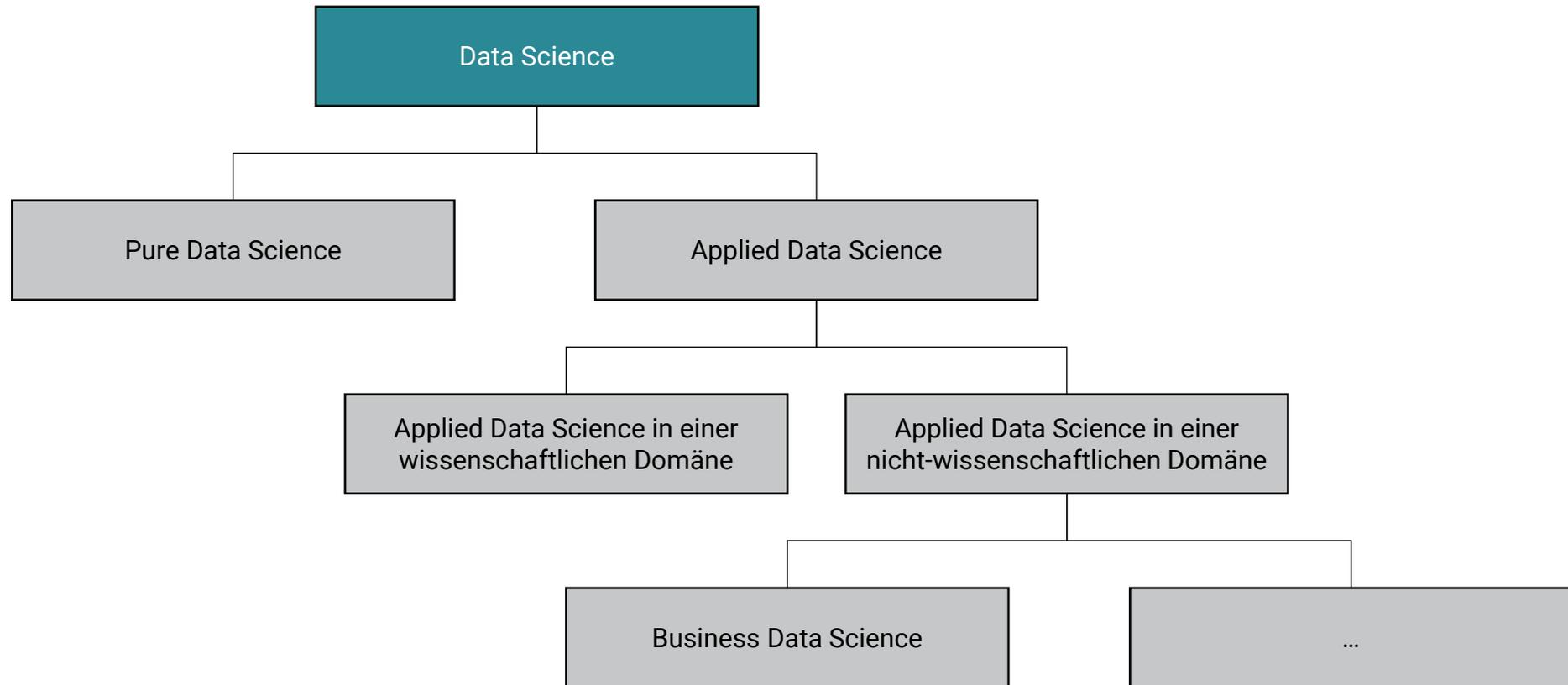


DASC-PM

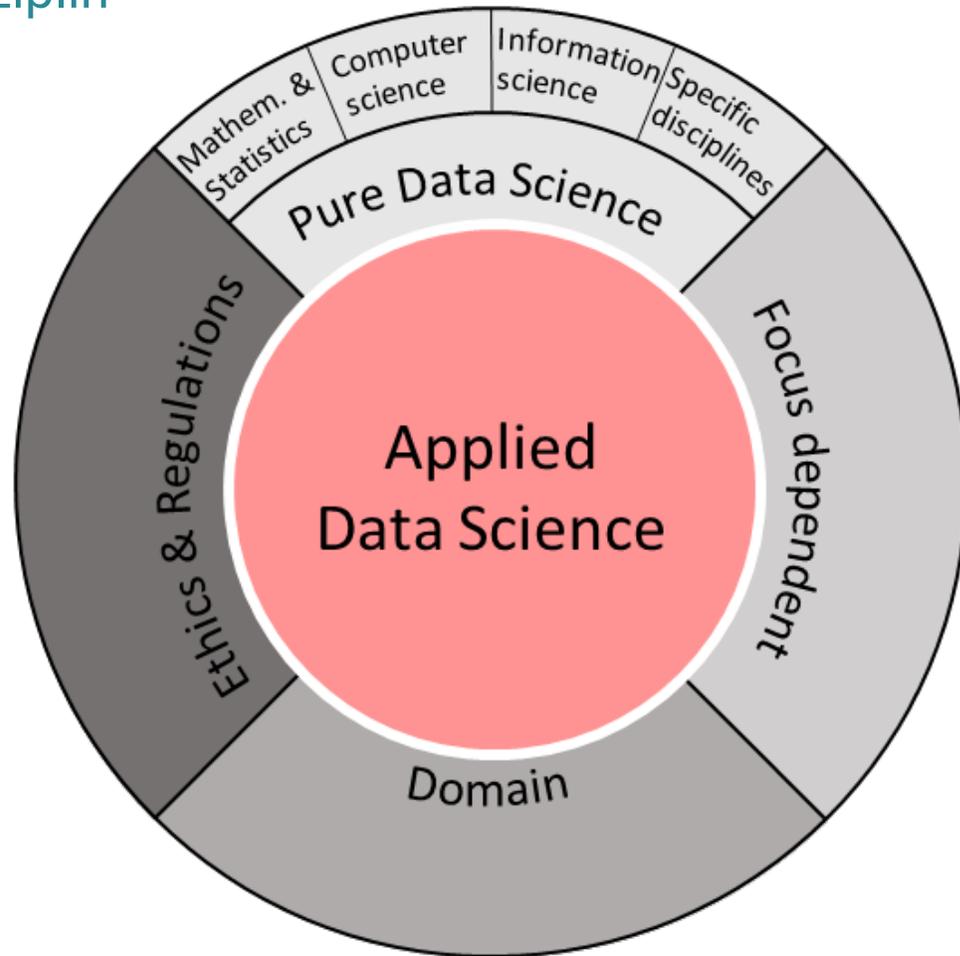
Ein Vorgehensmodell für
Data-Science-Projekte



Disziplin



Disziplin



Data Science ist ein interdisziplinäres Fachgebiet, in welchem mit Hilfe eines wissenschaftlichen Vorgehens, semiautomatisch und unter Anwendung bestehender oder zu entwickelnder Analyseverfahren Erkenntnisse aus teils komplexen Daten extrahiert und unter Berücksichtigung gesellschaftlicher Auswirkungen nutzbar gemacht werden.

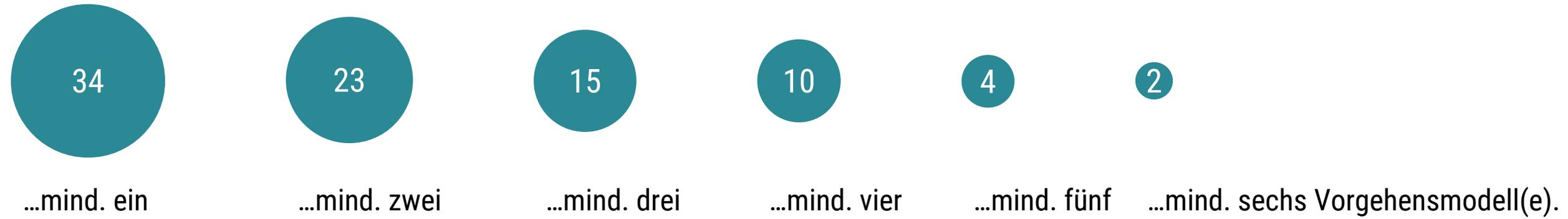
Morphologischer Kasten

Ziel / Ergebnis	Lösung einer konkreten Problemstellung		Erkenntnisgewinn		Forschungsbeitrag	
Datenbeschaffung	Alle Daten verfügbar und direkt nutzbar		Datenaufbereitung erforderlich	Zusammenführung unterschiedlicher Datenbestände	Neue Datenerhebung erforderlich	
Neuheitsgrad der verwendeten Lösungsverfahren	Nutzung von Standardverfahren		Nutzung angepasster Verfahren	Weiterentwicklung bestehender Verfahren	Neuentwicklung erforderlich	
Wiederverwendbarkeit	Nur projektbezogen	Unmittelbare Übertragbarkeit	Übertragbarkeit mit Anpassungen	Generalisierbarkeit der Ergebnisse	Generalisierbarkeit des Analyseverfahrens	
Potenzielle Felder von Unklarheiten	Domäne	Daten	Analyseverfahren	Nutzbarmachung	Nutzung	IT-Infrastruktur
Besondere Ressourcenanforderungen	Finanzen		Sachmittel		Personalressourcen (Anzahl / Kompetenz)	Zeit
Rollen außerhalb des Data-Science-Teams	Auftraggeber	Datenbereitsteller	Umsetzer	Ergebnisempfänger	Datenempfänger	Analyseverfahrensempfänger

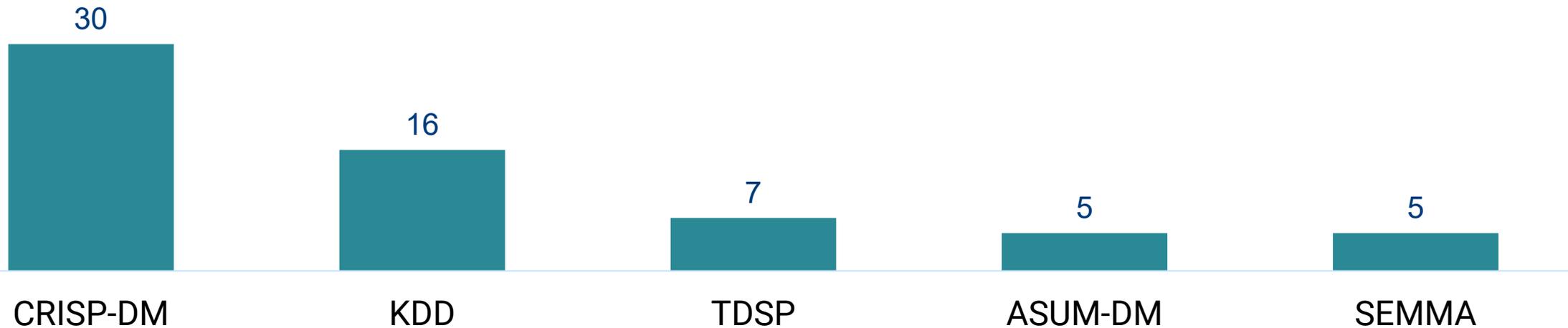
- Mangelhafte Datenqualität
- Unklare Zielsetzung
- Unzureichende Expertise des Projektteams
- Mangelhafte Zusammenarbeit im Team
- Zeit- und Ressourcenmangel
- Mangelhafte Akzeptanz und Integration der erzielten Ergebnisse

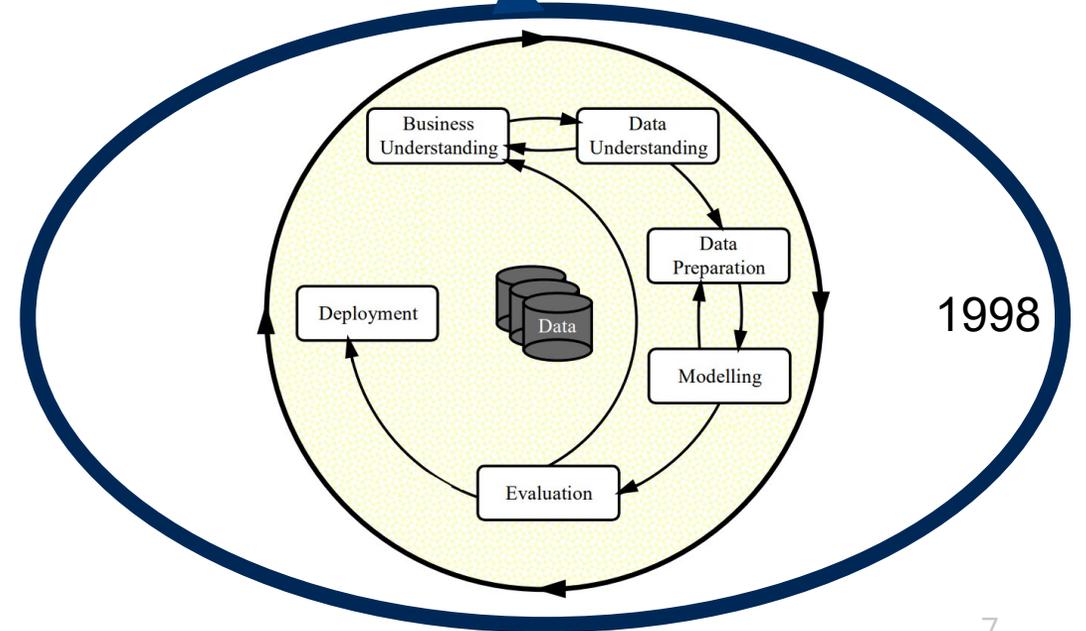
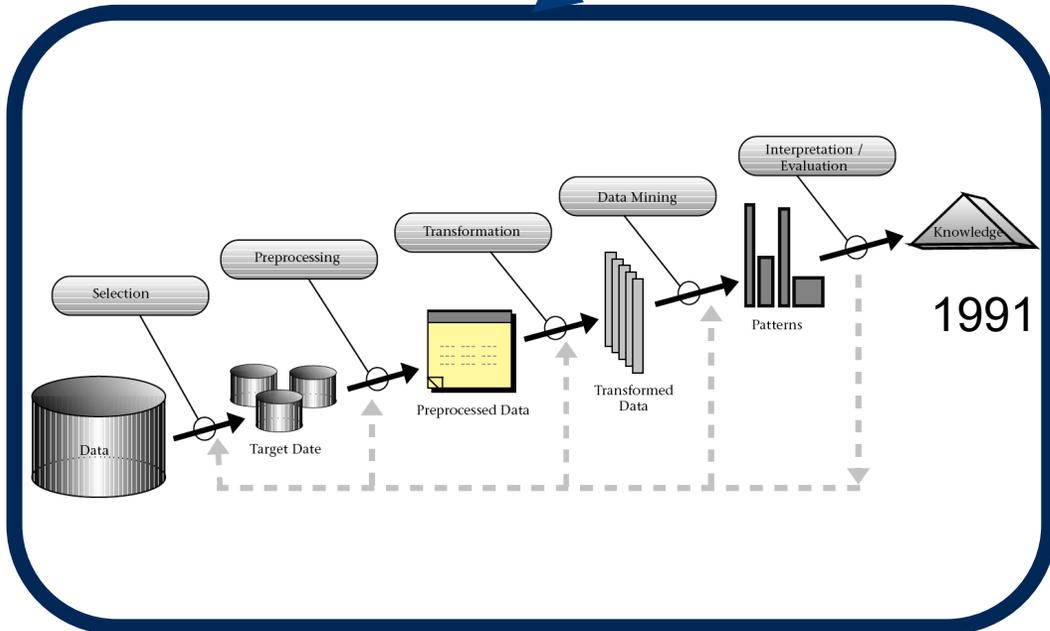
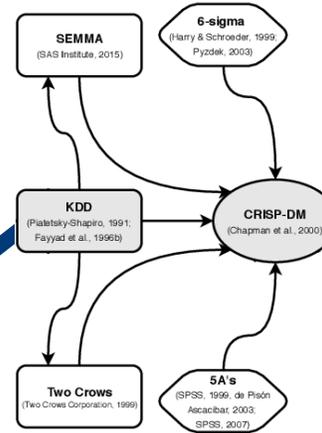


Von 40 TeilnehmerInnen kennen...

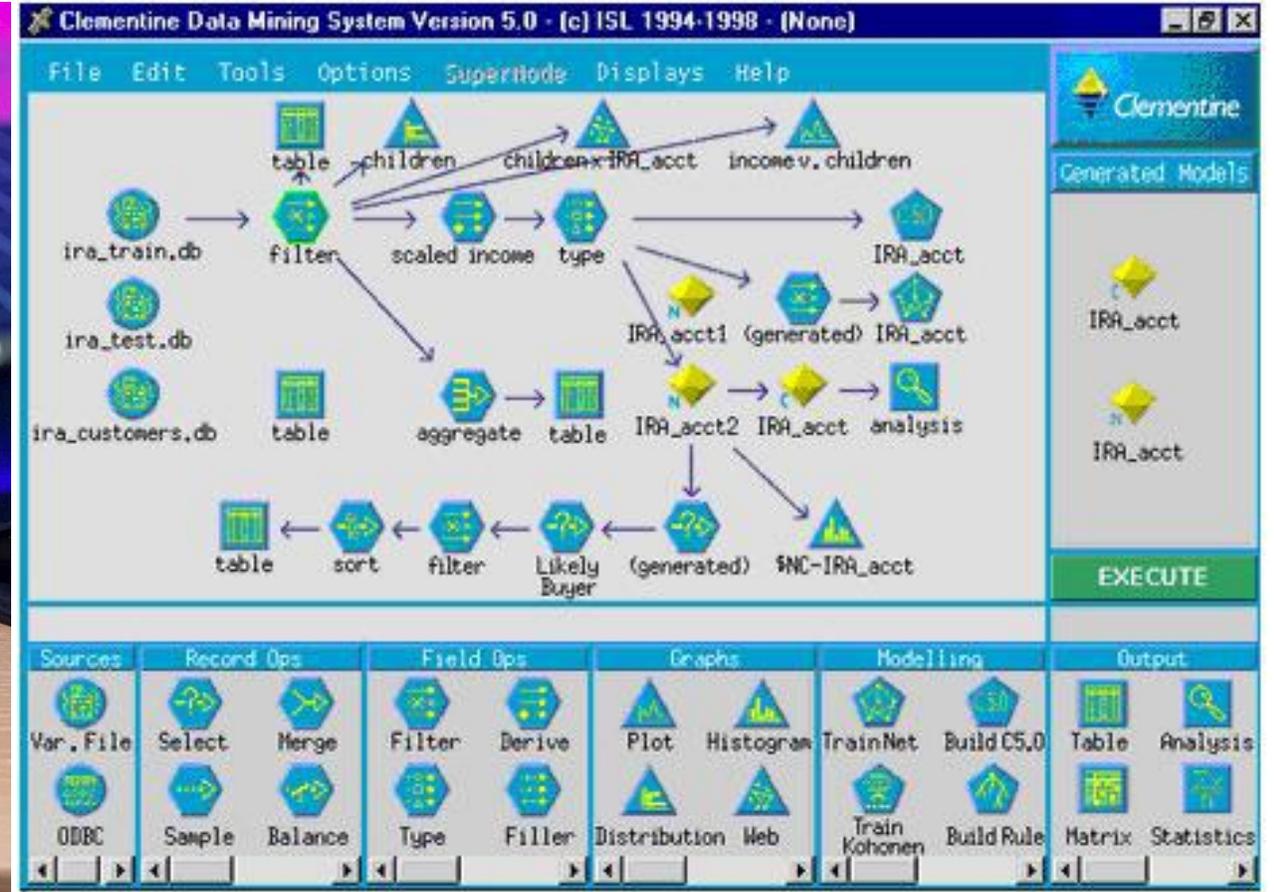


Bekannteste Vorgehensmodelle





Software für Data Science (Data Mining) im Jahr 1998



https://www.reddit.com/r/thinkpad/comments/kr7wjk/more_of_my_thinkpad_365xd_nt_40/

<https://webdocs.cs.ualberta.ca/~zaiane/courses/cmpu690/work/group5/clementine.html>

Home	Project Overview	Process Model	Project Partners	Special Interest Group	News	Privacy
------	------------------	---------------	------------------	------------------------	------	---------

CRISP-2.0: Updating the Methodology

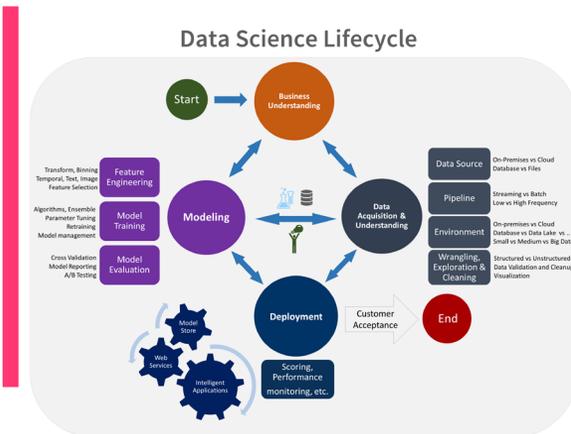
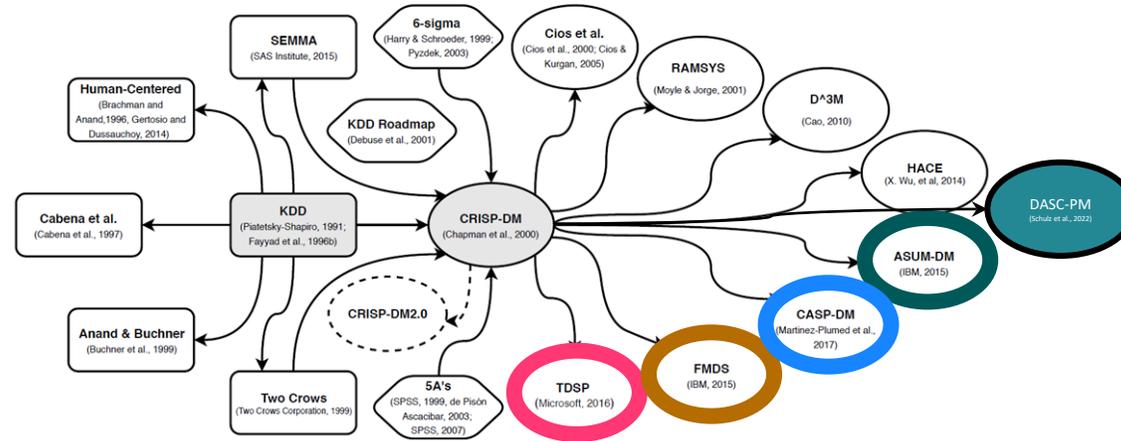
Why Update CRISP-DM?

CRISP-DM is a methodology that makes data mining and predictive analytics projects more efficient, better organized, more reproducible, more manageable, and more likely to yield business success.

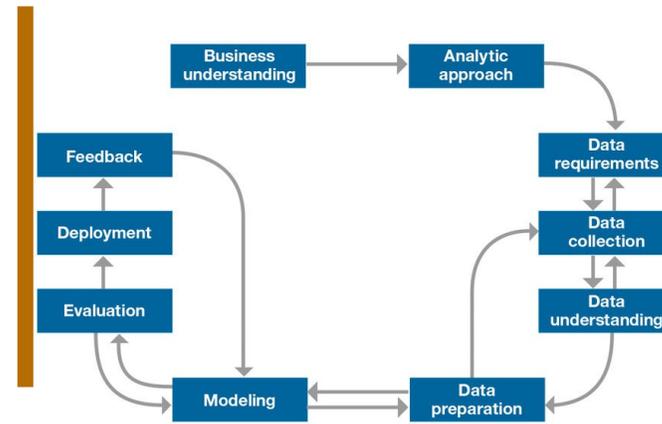
Many changes have occurred in the business application of data mining since CRISP-DM 1.0 was published. Emerging issues and requirements include:

- The availability of new types of data—text, Web, and attitudinal data, for example —along with new techniques for pre-processing, analyzing, and combining them with related case data
- Integration and deployment of results with operational systems such as call centers and Web sites
- Far more demanding requirements for scalability and for deployment into real-time environments
- The need to package analytical tasks for non-analytical end users and integrate these tasks in business workflows
- The need to seamlessly integrate the deployment of results and closed-loop feedback with existing business processes
- The need to mine large-scale databases in situ, rather than exporting an analytical dataset
- Organizations' increasing reliance on teams, making it important to educate greater numbers of people on the processes and best practices associated with data mining and predictive analytics

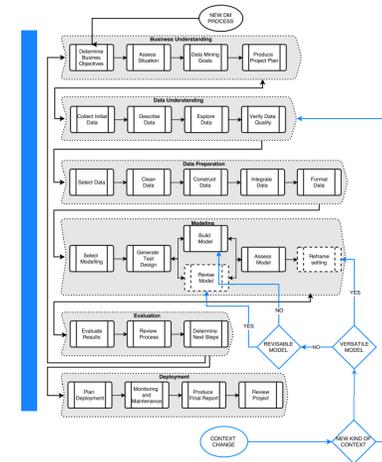
Join the CRISP-DM 2.0 Special Interest Group



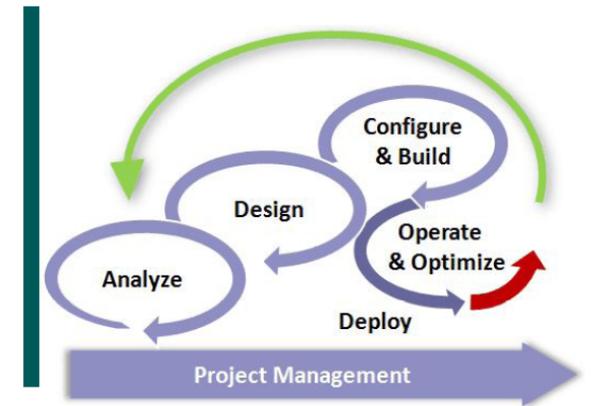
TDSP



FMDS



CASP-DM

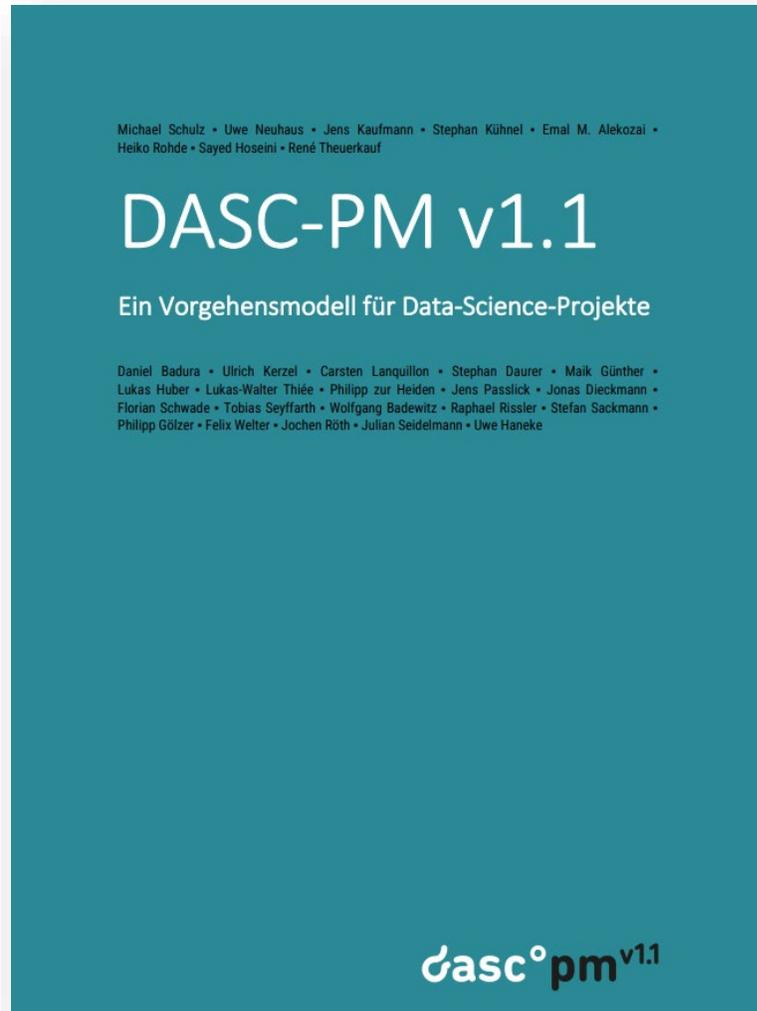


ASUM-DM

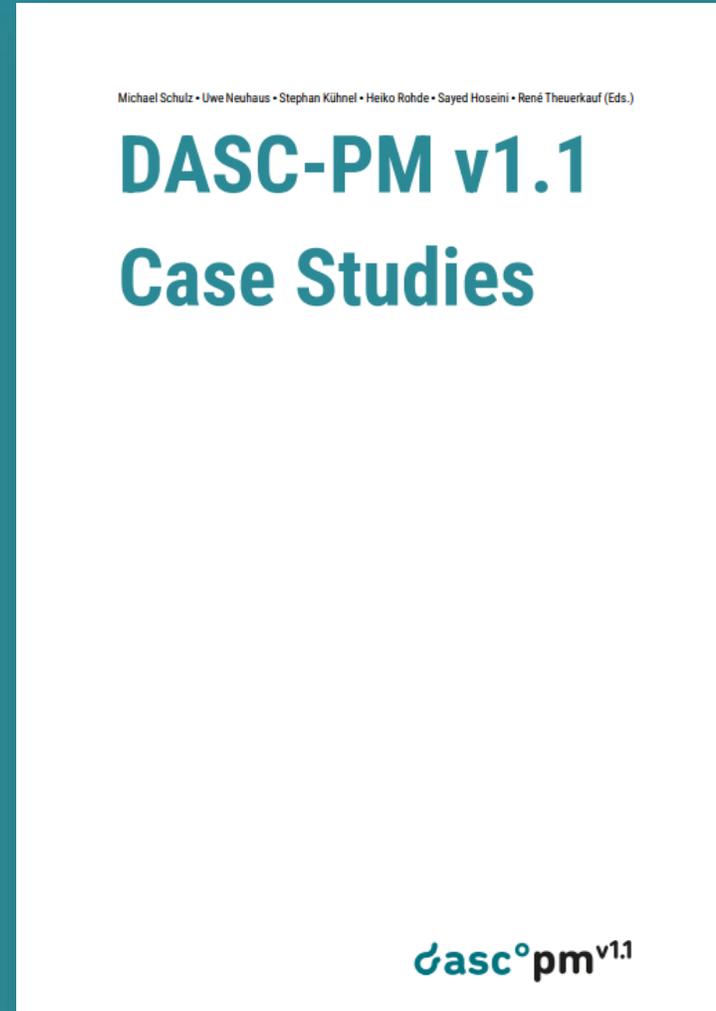
Aufruf zur Teilnahme an der Entwicklung eines Standardvorgehensmodells für Data-Science-Projekte

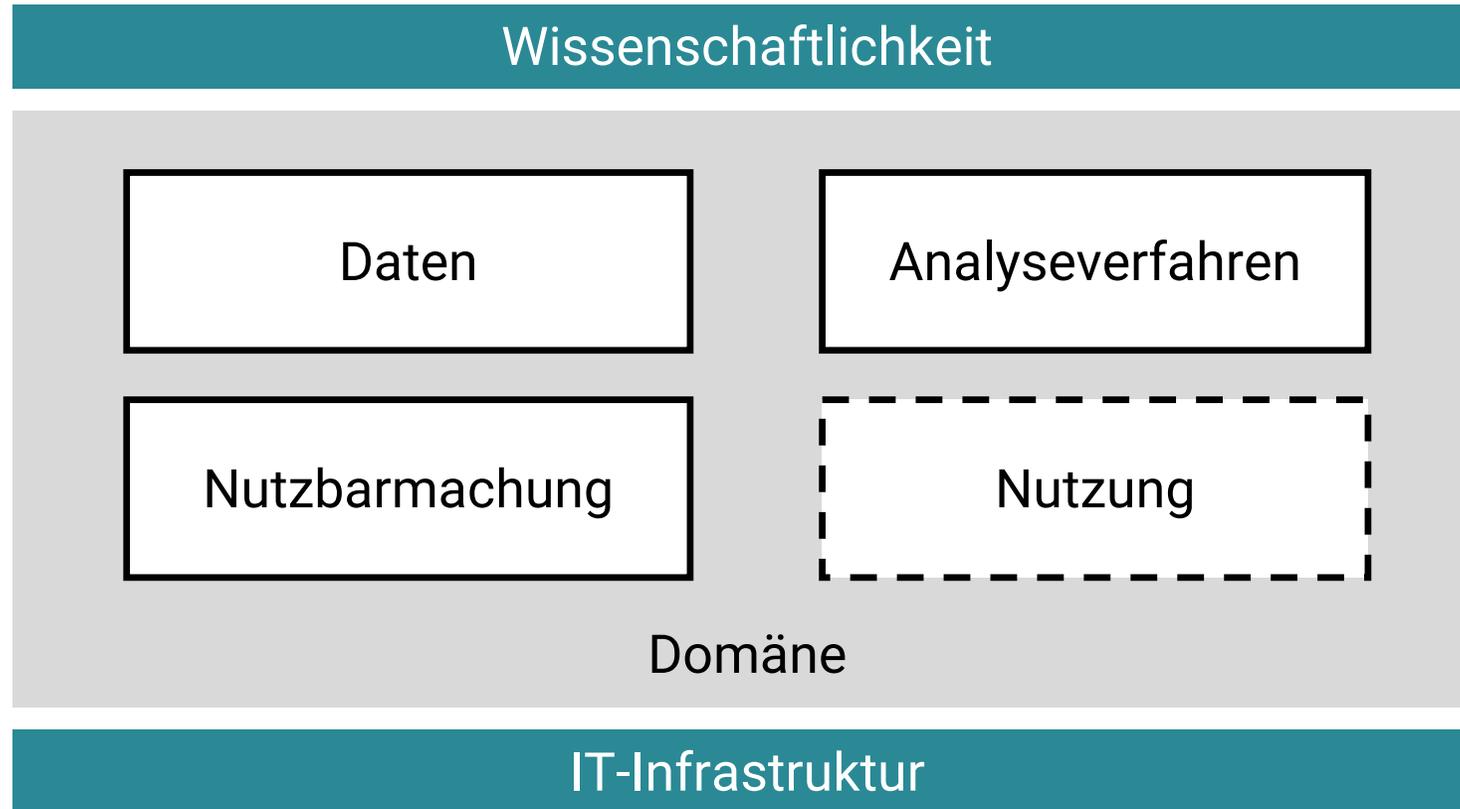
In einer virtuellen Arbeitsgruppe möchten wir zusammen mit Ihnen in den kommenden Monaten neben einer Definition des Data-Science-Begriffes vor allem ein Standardvorgehensmodell für Projekte in diesem Bereich entwickeln. Zielgruppe des Aufrufes sind Praktiker und Wissenschaftler, die im Data-Science-Bereich tätig sind. Durch die Zusammenarbeit in einer organisationsübergreifenden Arbeitsgruppe werden nicht nur heterogene Sichtweisen berücksichtigt, sondern auch verschiedene Erfahrungen gebündelt. Das Ergebnis soll Ihnen dabei helfen, eigene Data-Science-Projekte strukturiert und unter Einbeziehung sämtlicher Interessensgruppen erfolgreich durchführen zu können. Die Erkenntnisse sollen zudem unter Nennung aller beteiligten Personen veröffentlicht werden.

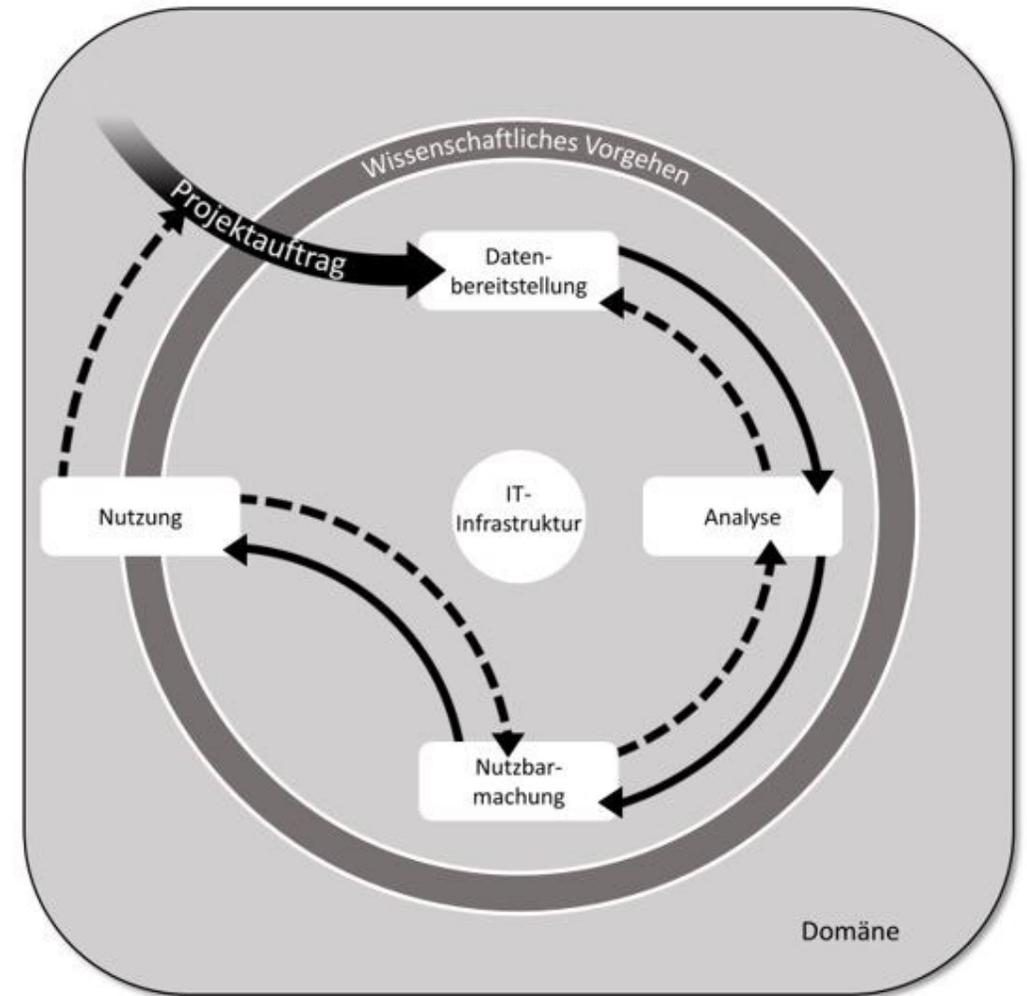
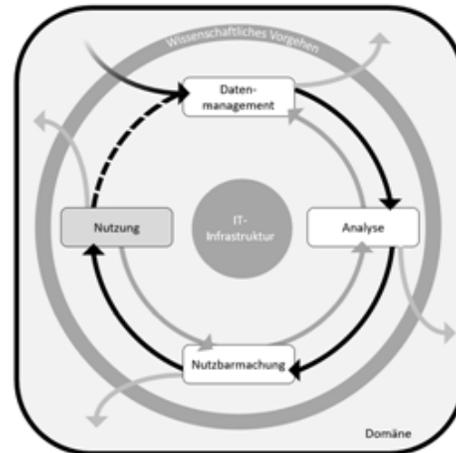
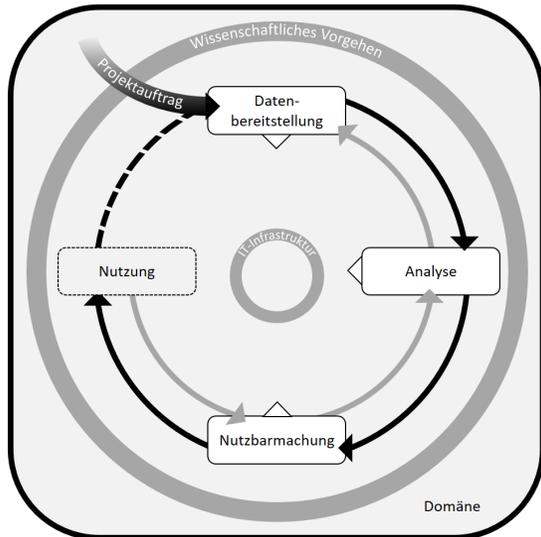
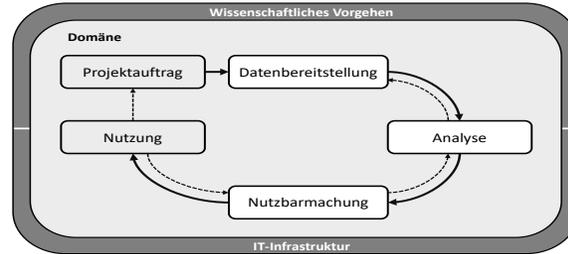
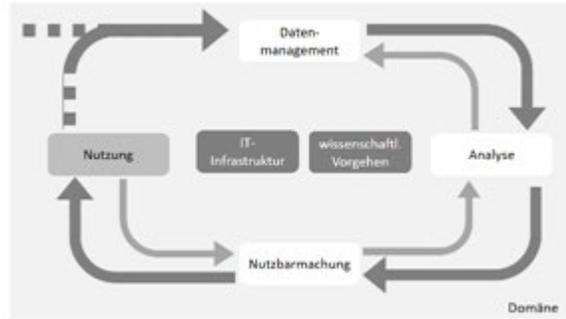
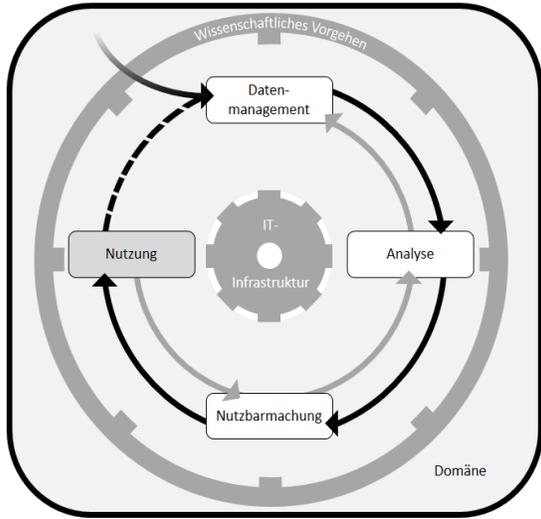
Interessenten können sich bis zum **31. März 2019** bei uns melden. Innerhalb eines halben Jahres möchten wir in zeitunabhängiger Zusammenarbeit und organisiert über verschiedene unabhängige Meilensteine gemeinsam mit Ihnen ein Ergebnis erarbeiten. Diese Aufteilung ermöglicht es Ihnen, sich in einem Maß zu beteiligen, das Sie an Ihre Arbeitsbelastung und Ihre Interessenschwerpunkte anpassen können. Die genaue Form der Zusammenarbeit wird von der Anzahl, der geografischen Verteilung und den fachlichen Hintergründen der Teilnehmer abhängig gemacht.

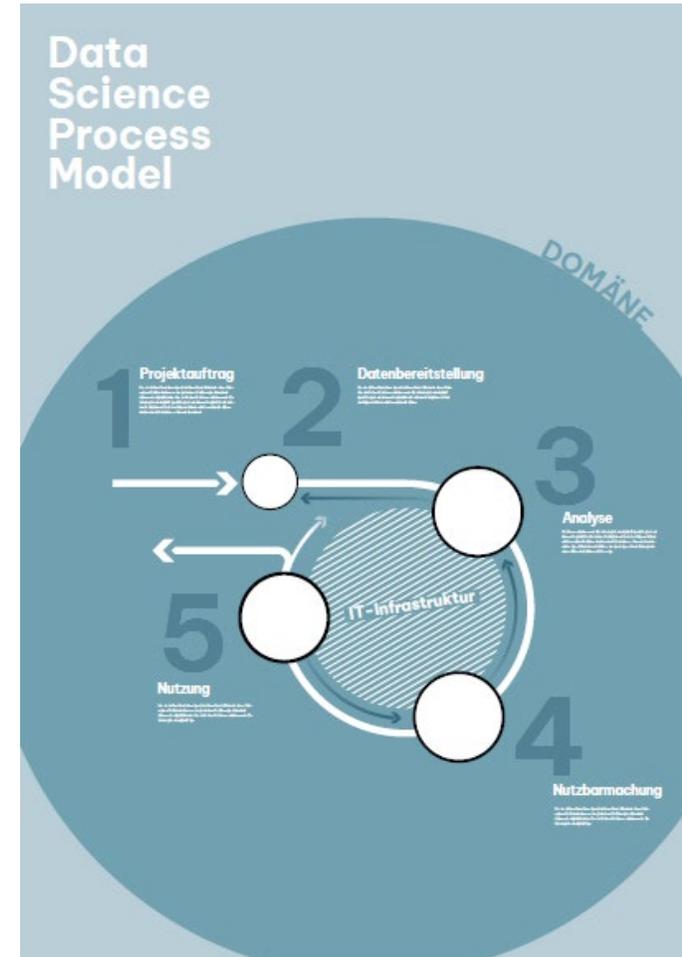
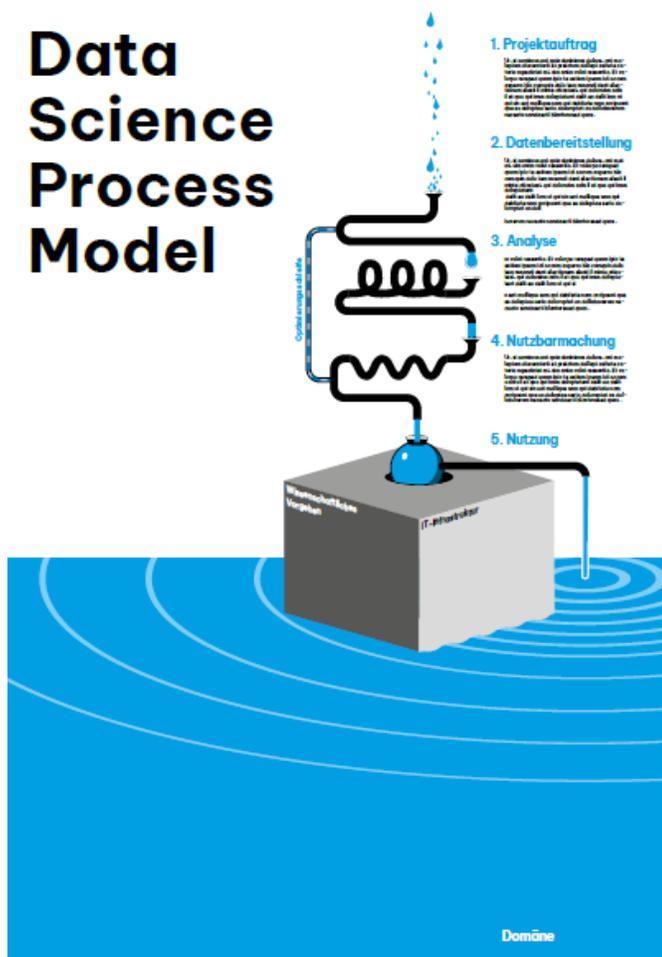


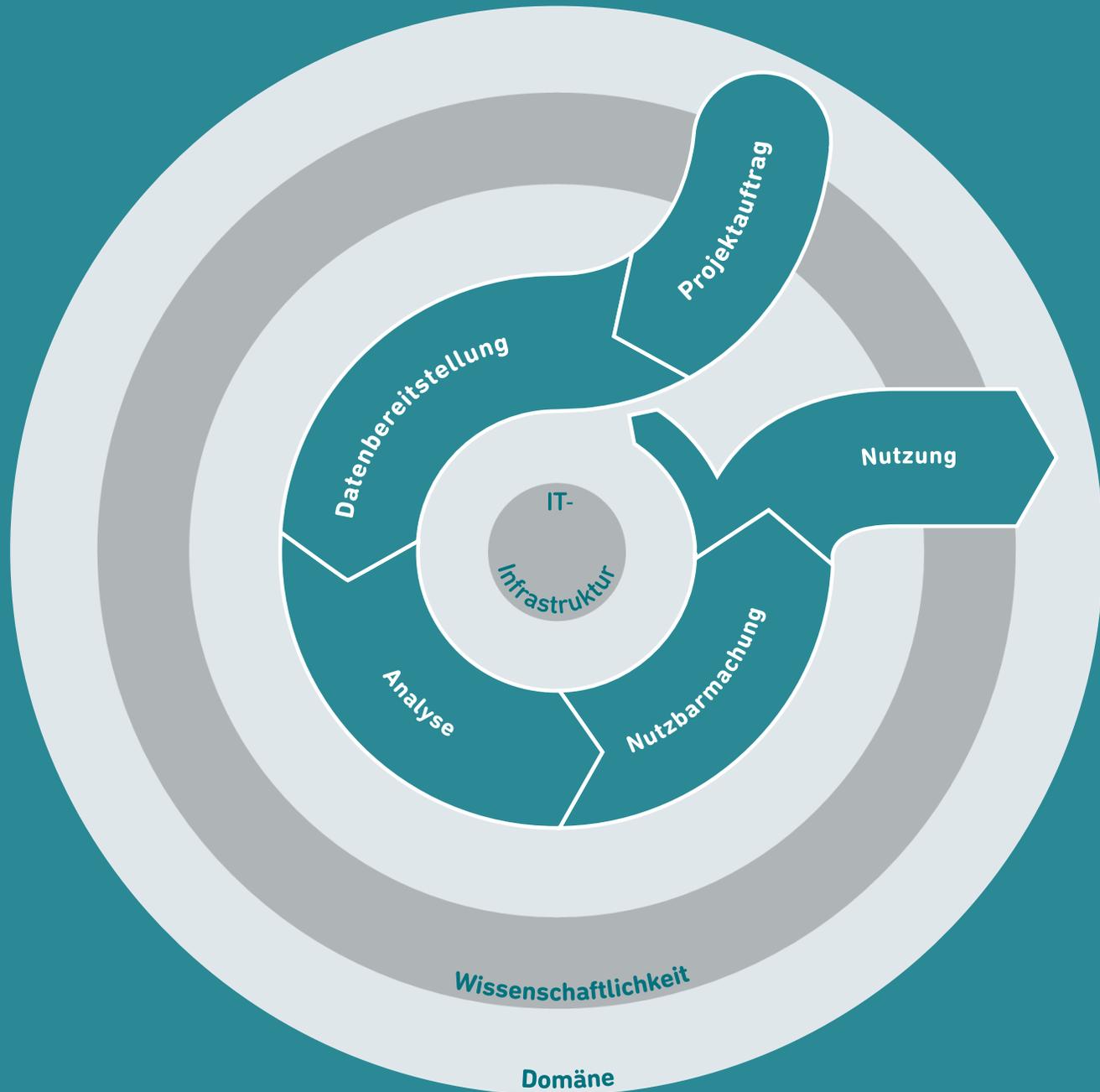
- Offene Arbeitsgruppe mit 60+ Mitgliedern (28 Autoren in der aktuellen Version)
- Veröffentlichung der Version 1.0 im Februar 2020, der Version 1.1 im März 2022
- Veröffentlichung einer englischsprachigen Version im Juni 2022
- Lizenziert unter der Creative Commons Attribution 4.0 International License.
- 108 Seiten





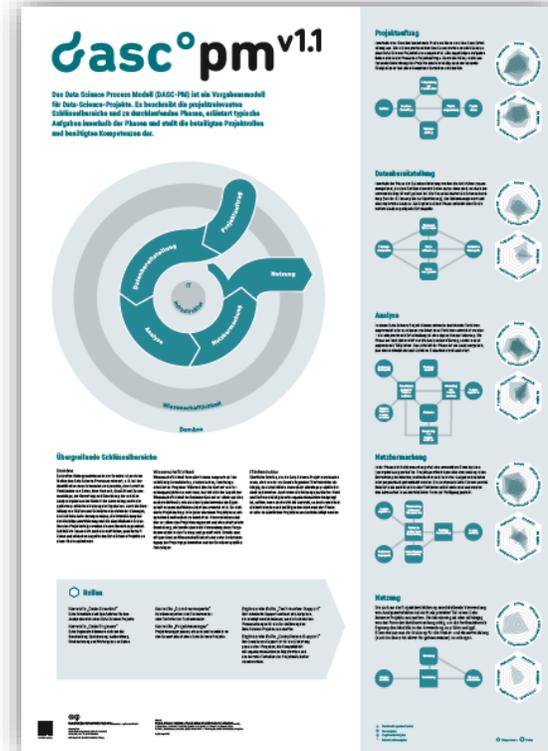








Das Vorgehensmodell als PDF (deutsch / englisch) kann frei heruntergeladen werden unter www.dasc-pm.org



Eine Poster-Darstellung findet sich im PDF des Vorgehensmodells. Für Druckversionen oder eine Einzelgrafik wenden Sie sich bitte an info@dasc-pm.org



Weitere Materialien senden wir Ihnen gerne zu. Bitte wenden Sie sich an info@dasc-pm.org

Data Scientist: *The Sexiest Job of the 21st Century*

Meet the people who can coax treasure out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

W

hen Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

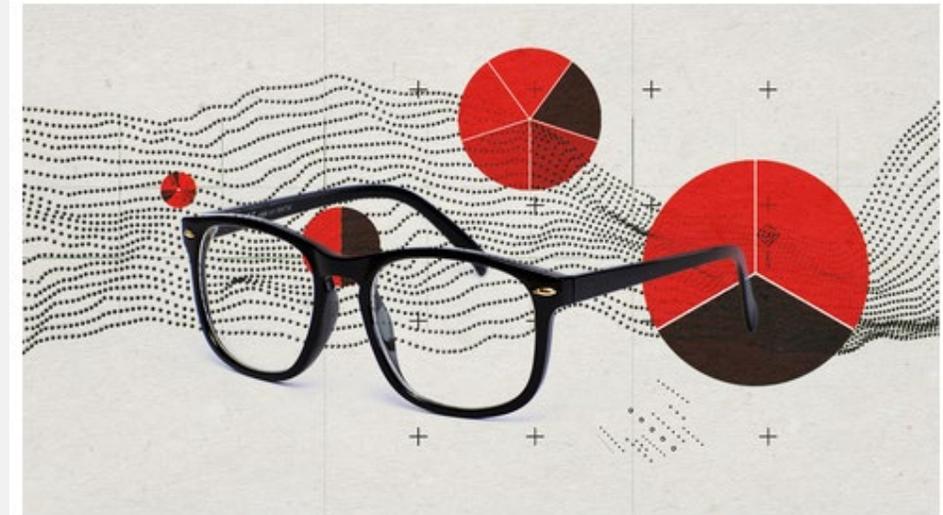
70 Harvard Business Review October 2012

2012

Is Data Scientist Still the Sexiest Job of the 21st Century?

by Thomas H. Davenport and DJ Patil

July 15, 2022



HBR Staff/StudioM1/Moritz Otto/Getty Images

2022

“Die vielen Führungskräfte, die erkannt haben, dass Data Science für ihr Unternehmen wichtig ist, müssen nun vielfältige Data-Science-Teams aufbauen und betreuen, anstatt nach Data-Scientist-Einhörnern zu suchen.”

Is Data Scientist Still the Sexiest Job of the 21st Century?

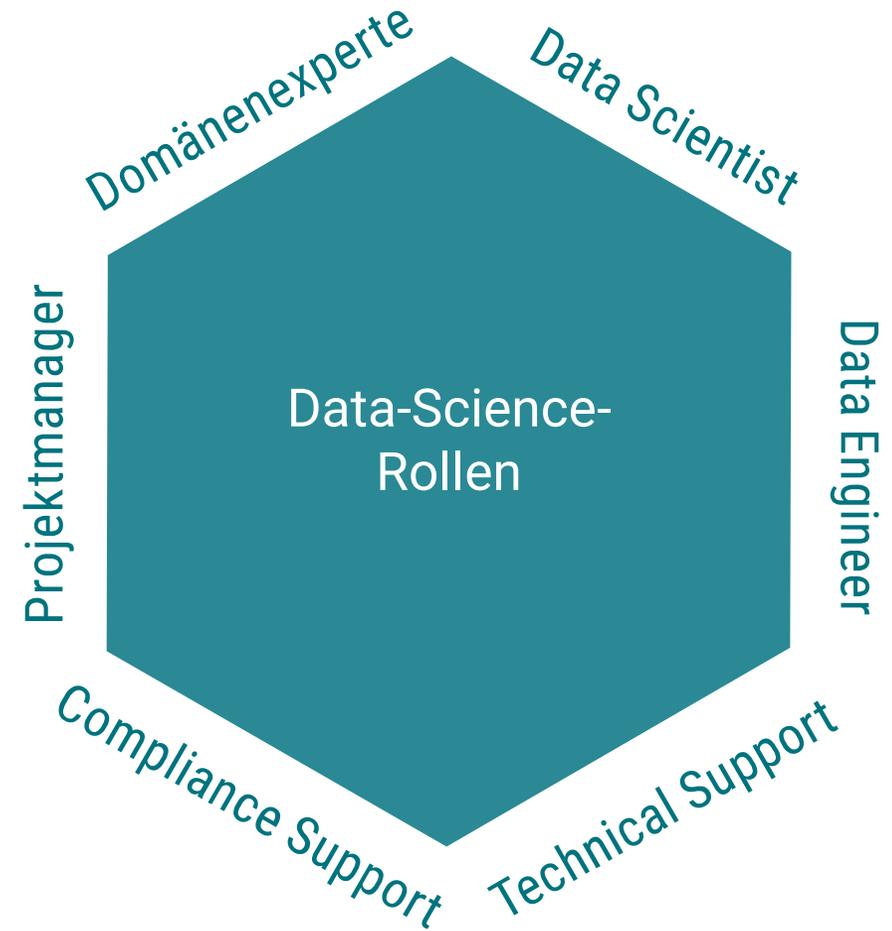
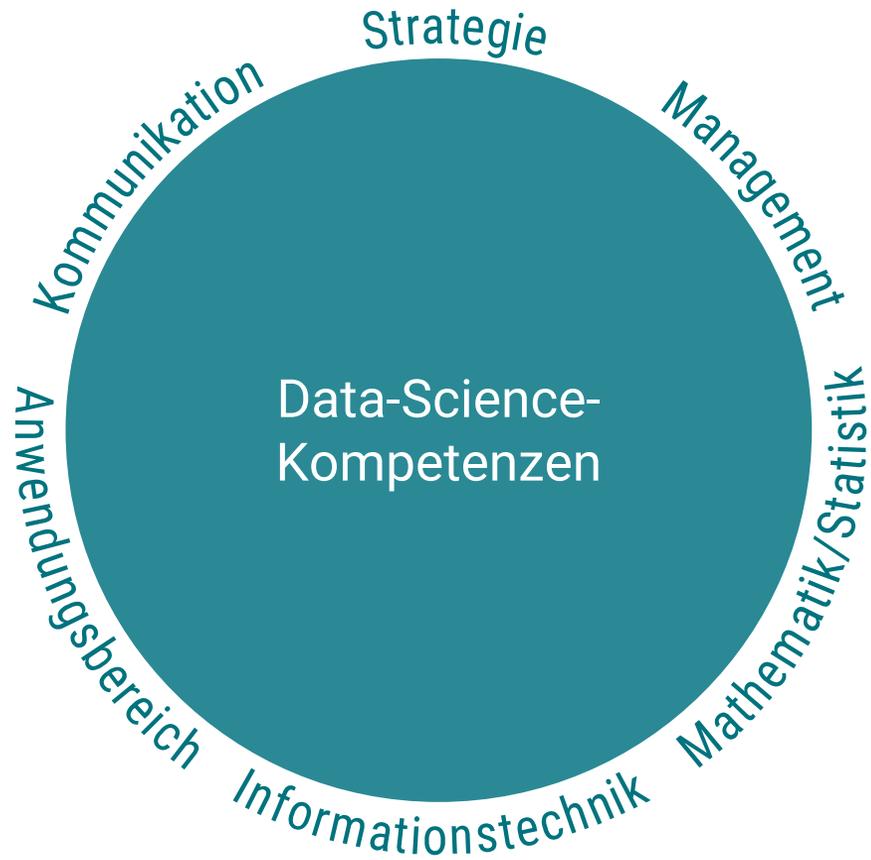
by Thomas H. Davenport and DJ Patil

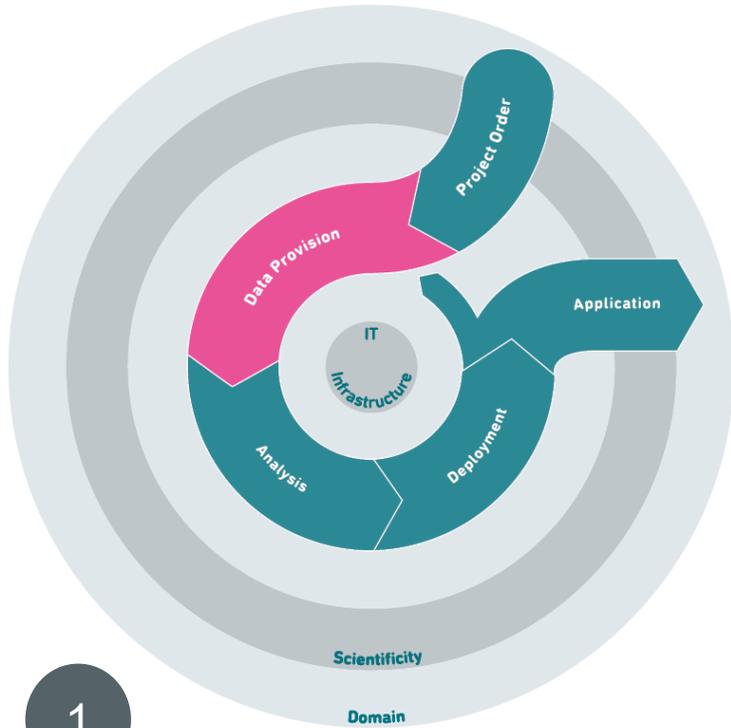
July 15, 2022



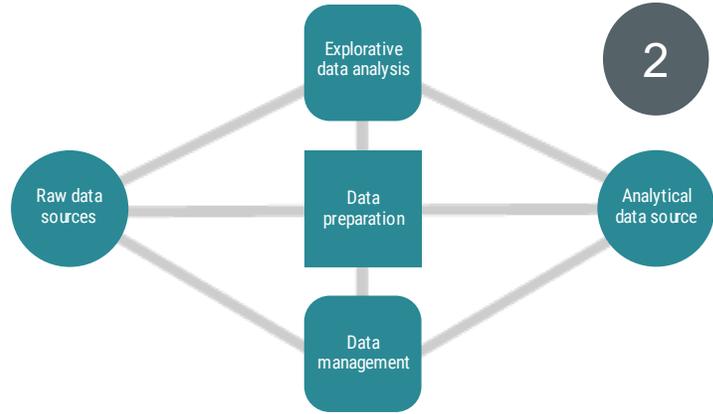
HBR Staff/StudioM1/Moritz Otto/Getty Images

2022

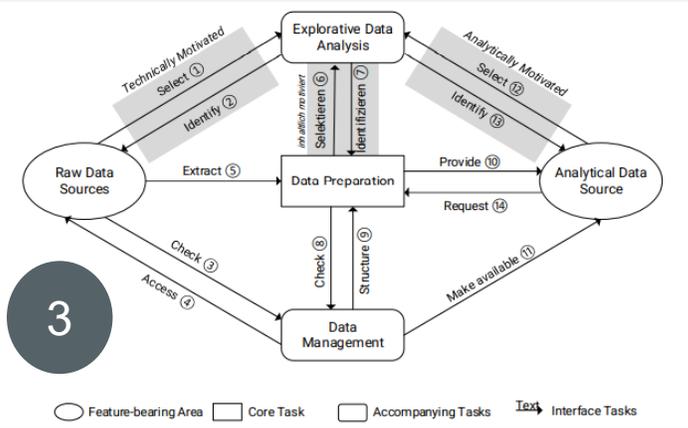




1



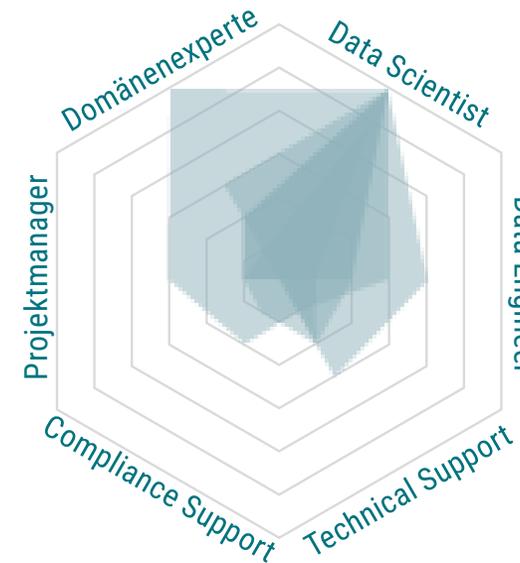
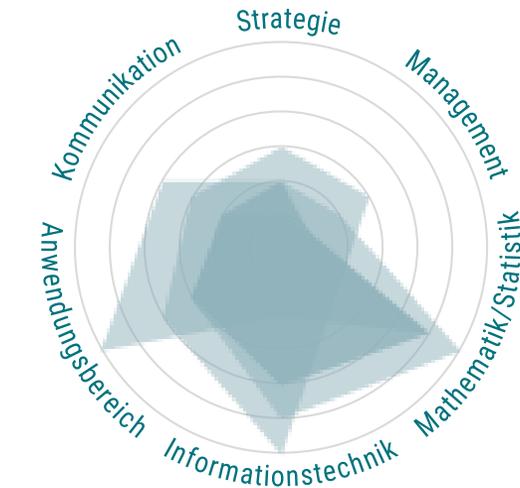
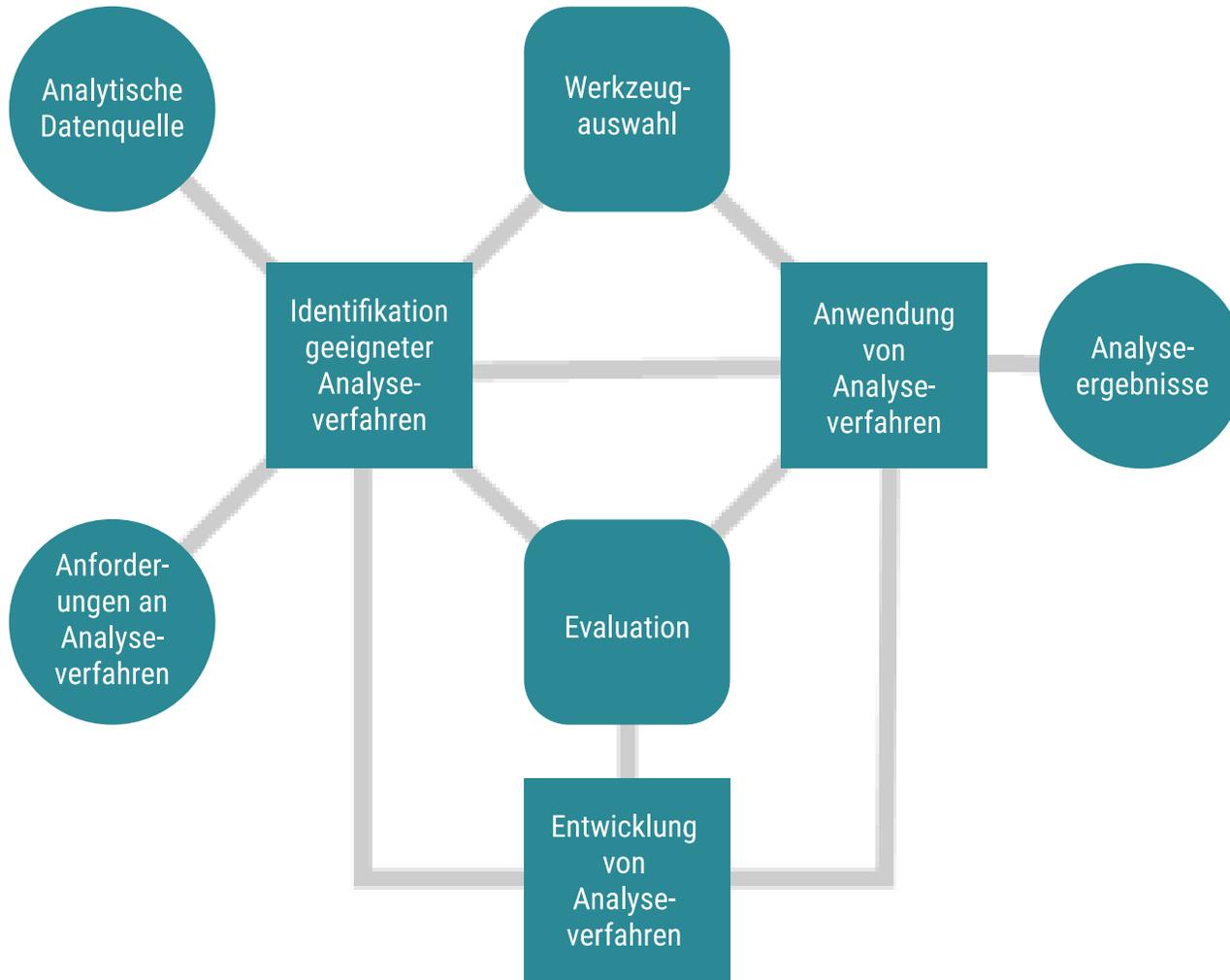
2



3

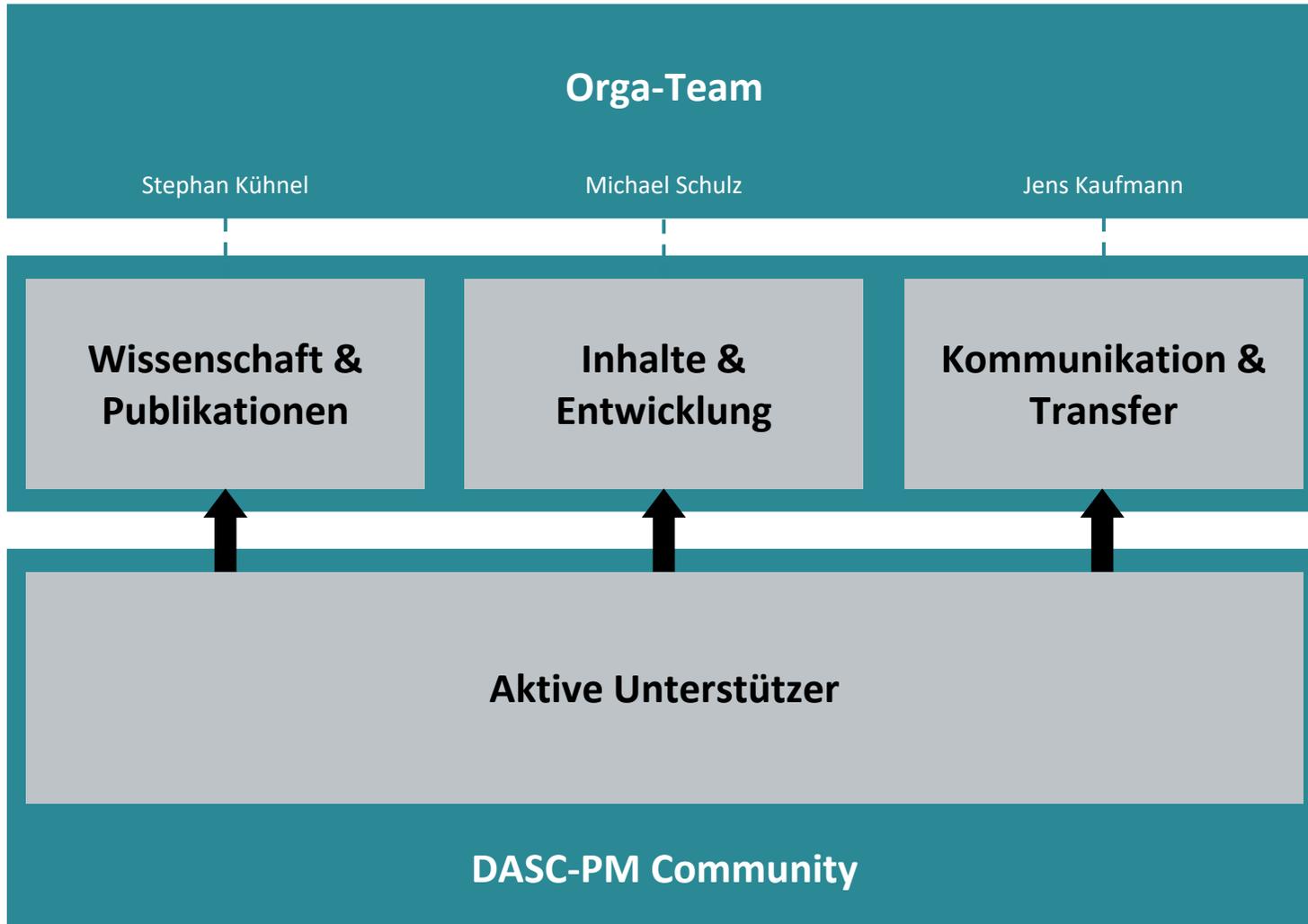
○ Feature-bearing Area □ Core Task □ Accompanying Tasks Text Interface Tasks

4



Vergleich von Vorgehensmodellen

	KDD <small>Fayyad et al., 1996</small>	CRISP-DM <small>Chapman et al., 2000</small>	ASUM <small>IBM Corporation, 2016</small>	ILG <small>Kienler, 2019a; Kienler, 2019b</small>	EDDA <small>Heserius et al., 2019</small>	TDSP <small>Microsoft, 2020</small>	DASC-PM <small>Schulz et al., 2020</small>	
Objective and Economy	●	●	●	●	●	●	●	Goals and Requirements
Needs from User Perspective	○	●	○	○	○	○	○	
Analysis of Affected Processes	○	●	○	○	○	○	○	
Analysis of Key Activities	○	●	○	○	○	○	○	
Legal	○	●	○	●	●	○	●	
Requirements	○	●	●	●	●	●	●	
Data Access	●	●	○	●	●	●	●	Structured Project Setup
Project Management	○	●	●	○	●	●	●	
Selection of Technology	○	●	○	●	○	●	●	
Project Team and Competences	○	●	●	●	●	●	●	
Data Preparation	●	●	○	●	●	●	●	Concepts and Implementation
Data Understanding and Exploration	●	●	○	●	●	●	●	
Selection of Models	●	●	○	○	●	●	●	
Model Building	●	●	●	●	●	●	●	
Robustness and Model-Security	●	○	○	○	●	●	●	
System Architecture	○	○	○	●	●	●	●	
Data Architecture	●	○	○	●	○	●	○	
Evaluation	●	●	●	●	●	●	●	
Deployment	●	●	●	●	●	●	●	Utilisation of the Results
Qualification and Adjustment of the Job Profiles	○	○	○	○	○	○	●	
Process Integration	○	○	○	○	○	○	●	



- Koordination der Aktivitäten
- Kommunikation intern/extern
- Roadmap-Verantwortung

- Strukturierung der Aktivitäten in Schwerpunkten
- Sicherstellen von kontinuierlich engagierten Ansprechpartnern

- Regelmäßiger Jour fixe (4x/Jahr)
- Aktive Teilnahme an Aktivitäten
- Anregung von Projekten und auf Wunsch Übernahme von Themenverantwortung

- Newsletter / Teilnahme an Umfragen / ...

Sollten Sie sich an Aktivitäten rund um das DASC-PM beteiligen möchten, genügt eine formlose E-Mail an info@dasc-pm.org

