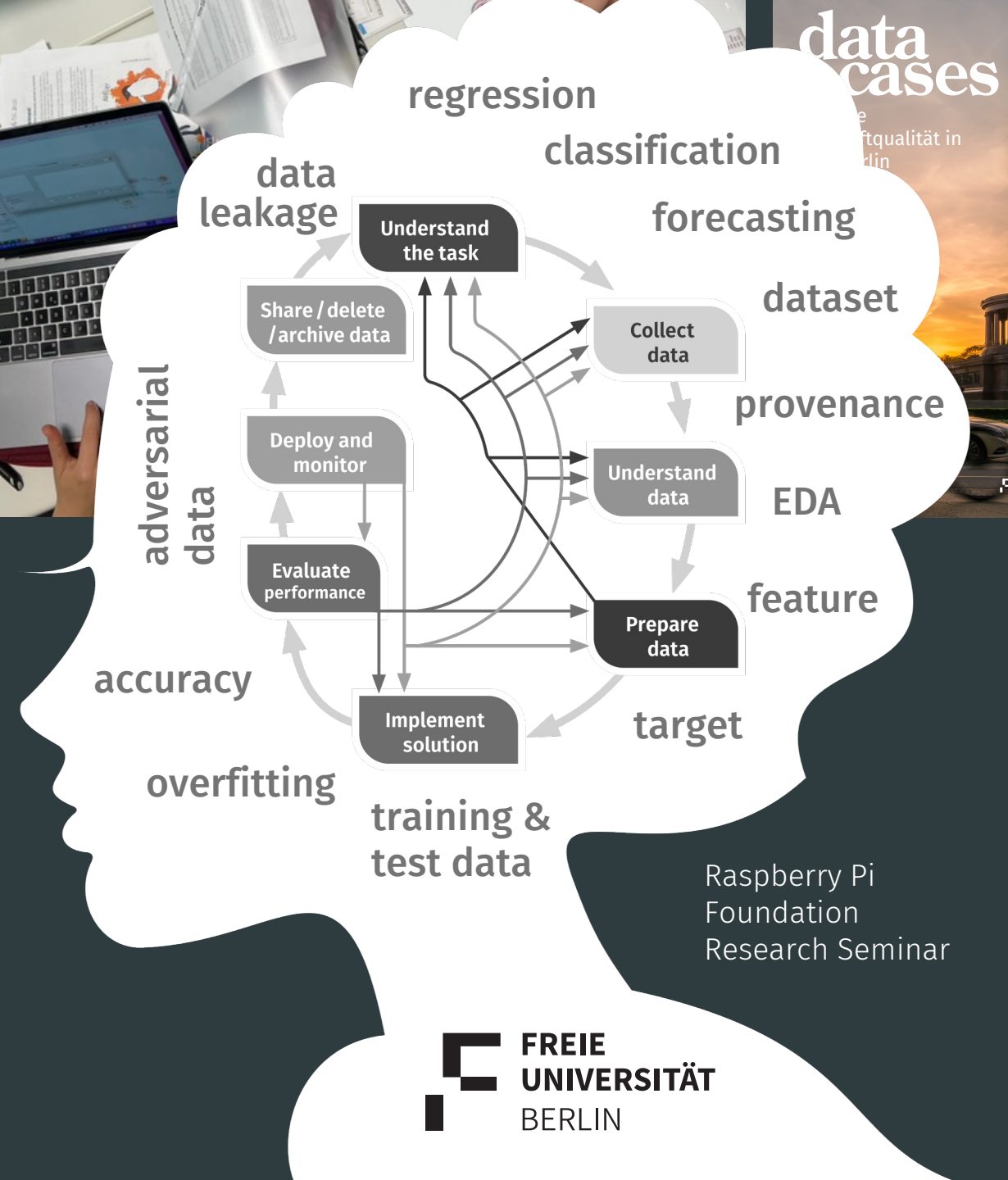




Data-related concepts and practices for AI education in K-12

Viktoriya Olari | 14.10.2025

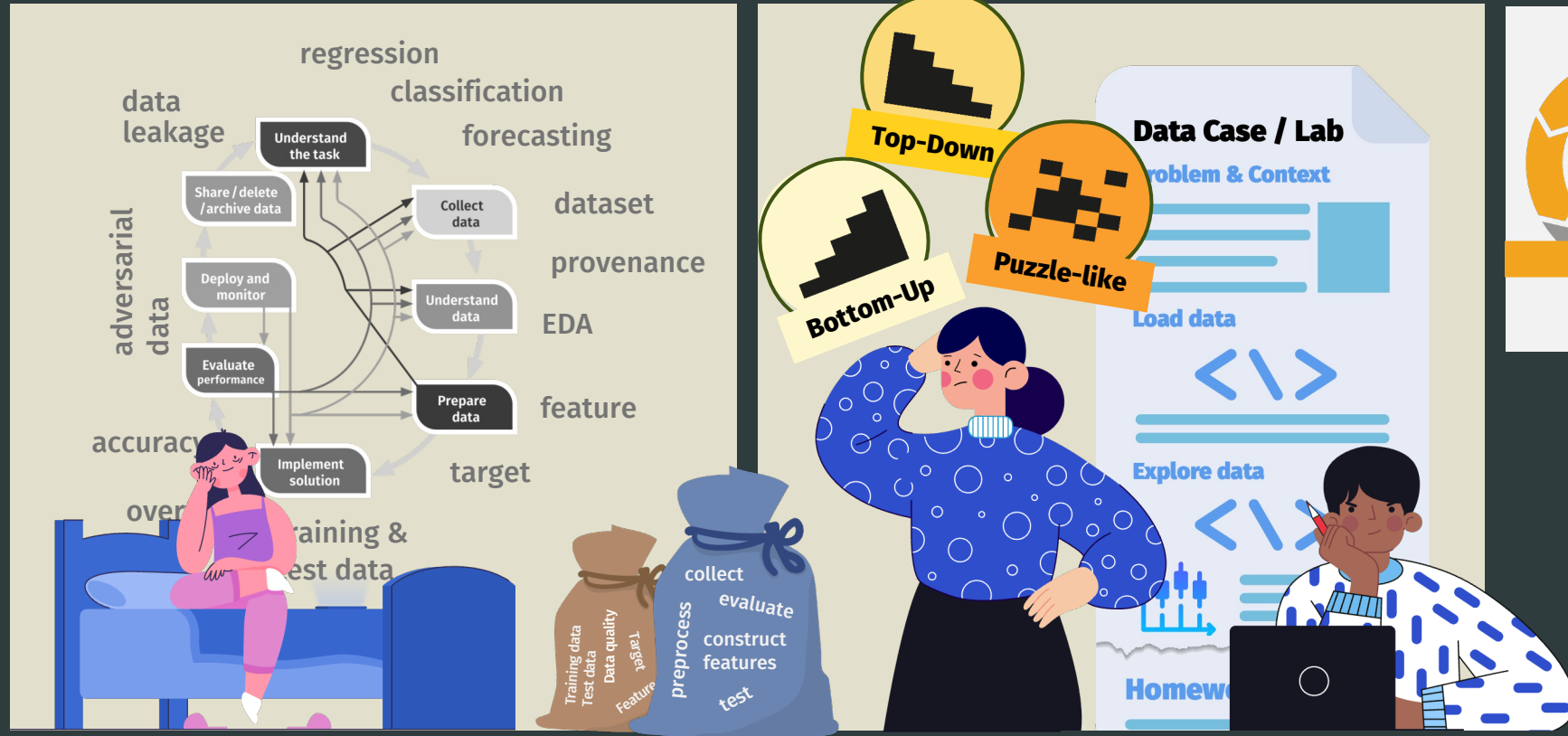


Roadmap

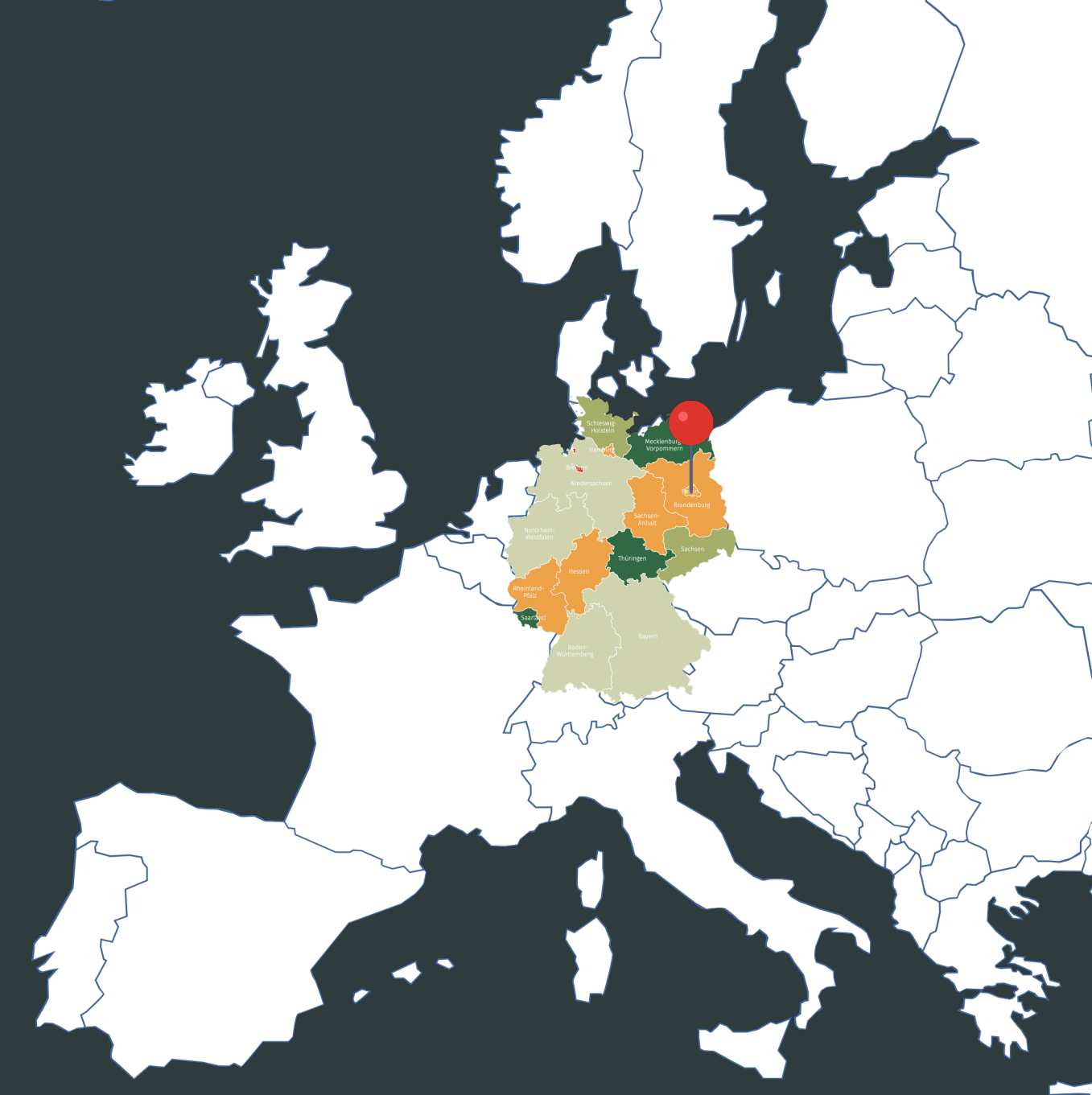
Data concepts & practices

Summary

Context & Motivation



The data case study



Context



Compulsory Computer Science Education starting in Grade 5

- At least **6 compulsory hours / week**
- At least **4 compulsory hours / week**
- At least **1 compulsory hour / week**
- Offered at individual school types or in individual grades
- No courses offered

Context

AI topics are entering formal school education.

Automaten und künstliche Intelligenz

Inhaltliche Schwerpunkte:

- Aufbau und Wirkungsweise einfacher Automaten
- Maschinelles Lernen mit Entscheidungsbäumen
- *Maschinelles Lernen mit neuronalen Netzen*

Die Schülerinnen und Schüler

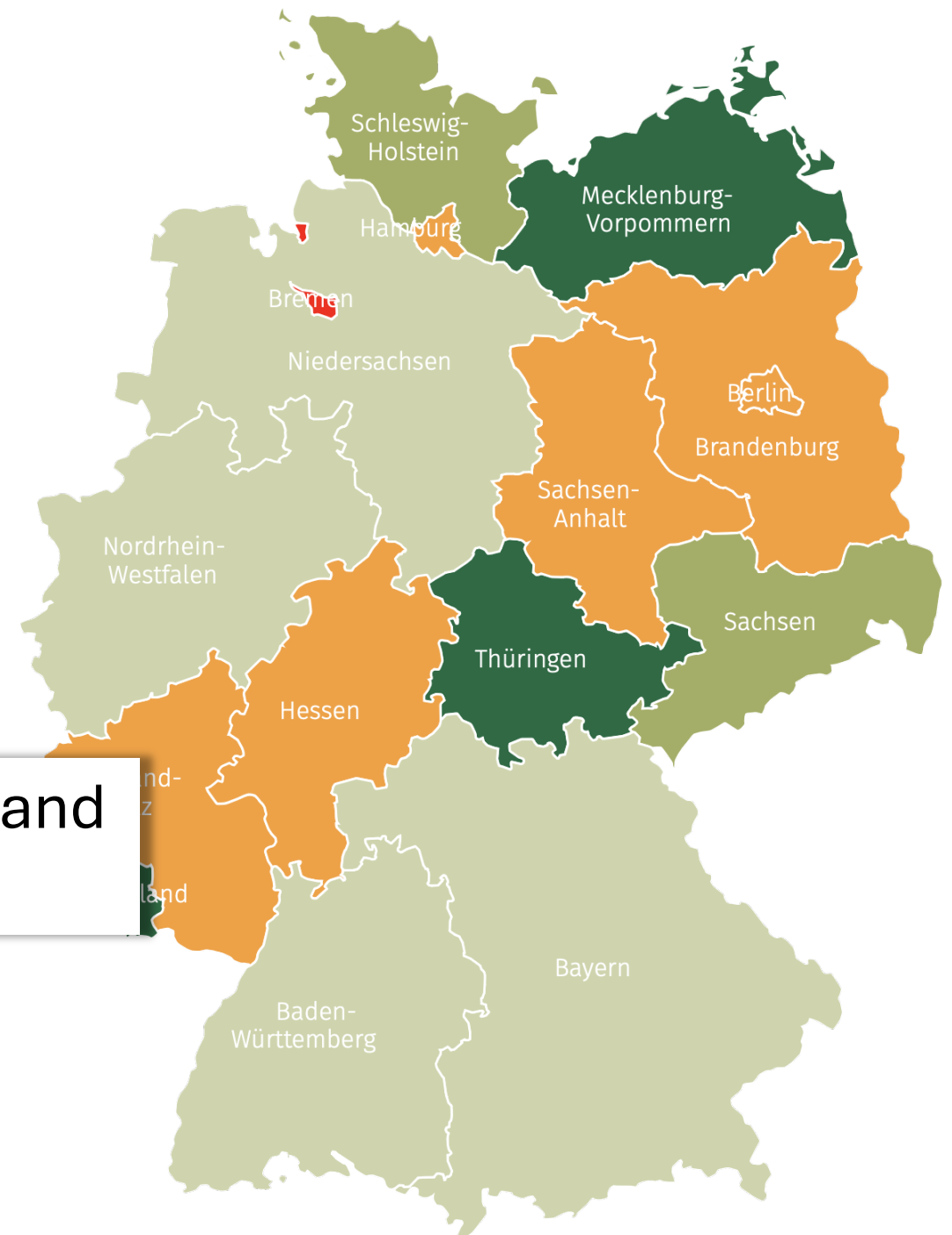
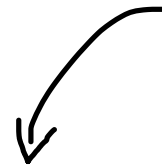
Machine Learning with decision trees and neuronal networks

des maschinellen Lernens dar (DI),

- *beschreiben die grundlegende Funktionsweise künstlicher neuronaler Netze in verschiedenen Anwendungsbeispielen (KK).*

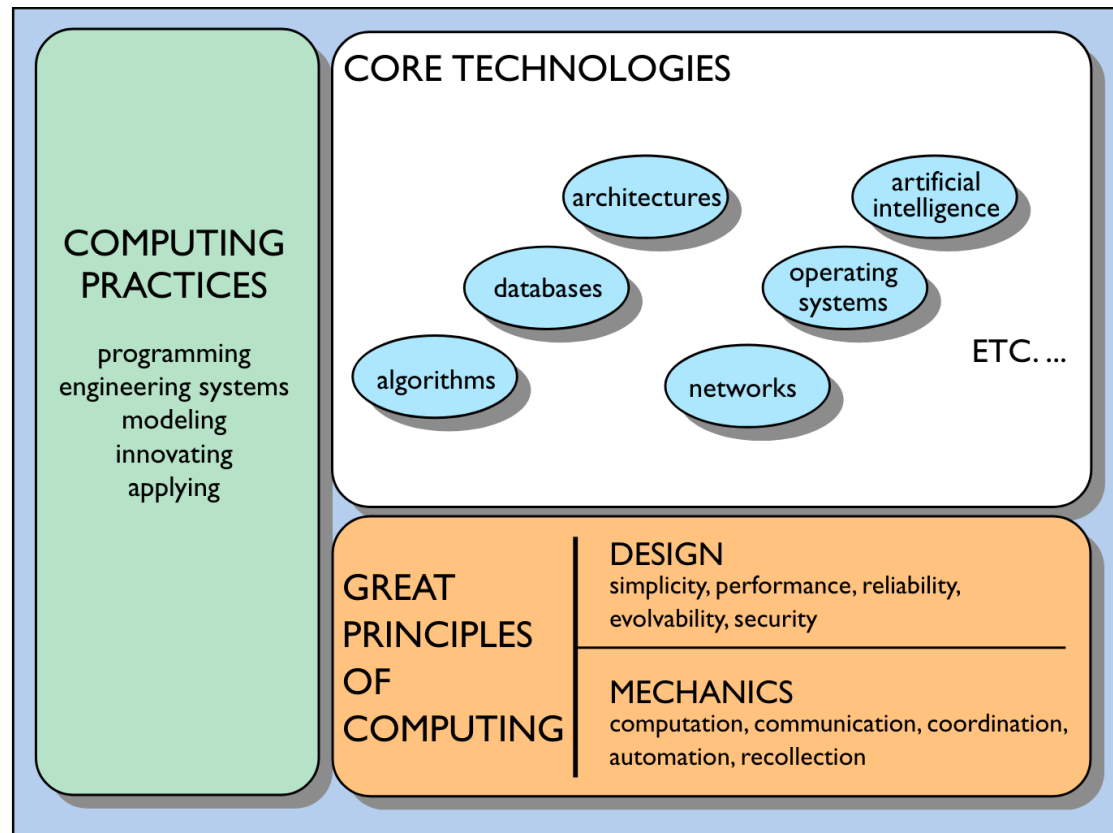
Core curriculum for secondary level I – grades 5 and 6 in North Rhine-Westphalia (2021)

for 5th
graders

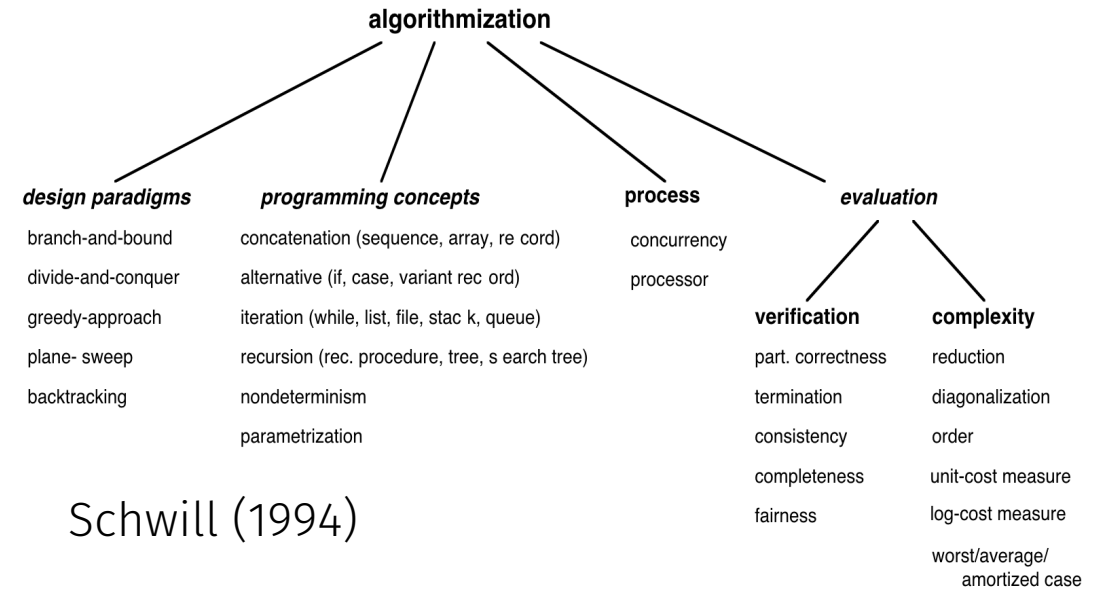


Motivation

Teaching should focus on key concepts and practices of CS.



Denning (2003)



Schwill (1994)



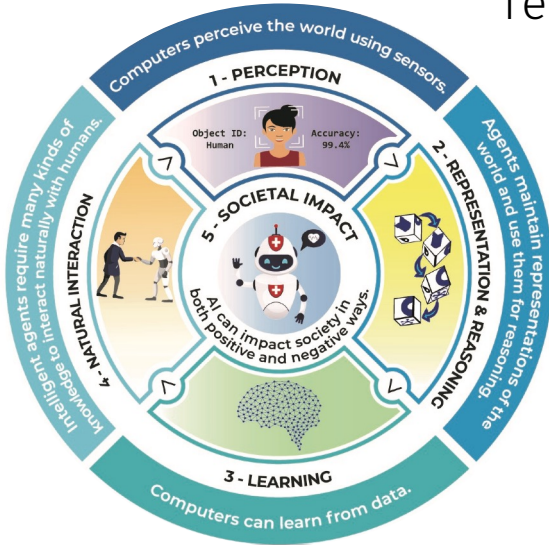
LEAD WITH CONCEPTS

Motivation

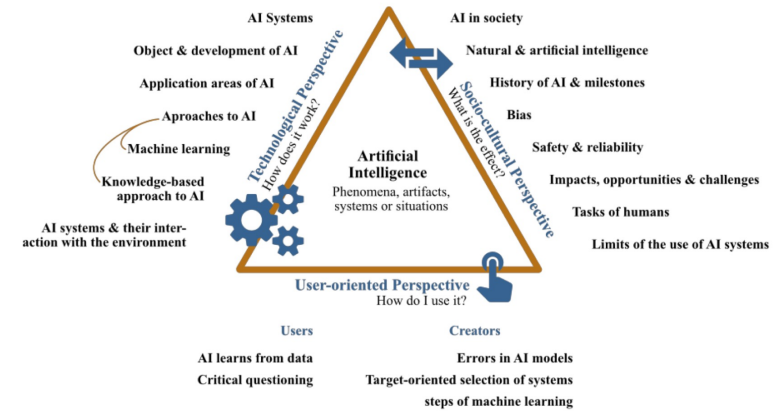
Conceptualisation is still ongoing!

CT 2.0

Tedre et al. (2021)



Touretzky et al. (2019)



Michaeli et al. (2022)

SEAME

Waite et al. (2023)



Conceptualisation is still ongoing!

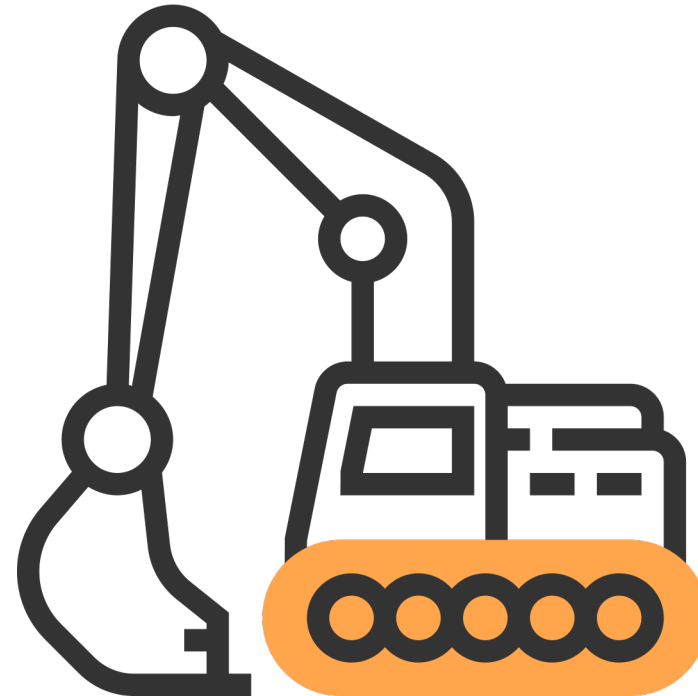
WiPSCE '21, October 18–20, 2021, Virtual Event, Germany

	Acquisition	Cleansing	Modeling	Implementation	Optimisation	Analysis	Visualisation	Evaluation	Sharing	Erasing	Archiving
[11]	-	-	✓	-	-	-	-	✓	-	-	-
[9]	✓	-	✓	-	-	✓	-	✓	-	-	✓
[1]	✓	-	-	-	-	-	-	✓	-	-	-
[14]	✓	-	-	-	-	-	-	✓	-	-	-
[3]	✓	-	-	-	-	✓	-	✓	-	-	-
[12]	✓	-	-	-	-	-	-	✓	-	-	-
[8]	-	-	-	-	-	-	-	-	-	-	-

Table 1: Coverage of the data literacy-related competencies in currently available AI educational frameworks, categorized by the stages of the data lifecycle proposed by [6].

The **role of data** is underestimated.

Olari & Romeike (2021)



Motivation

The key role of data

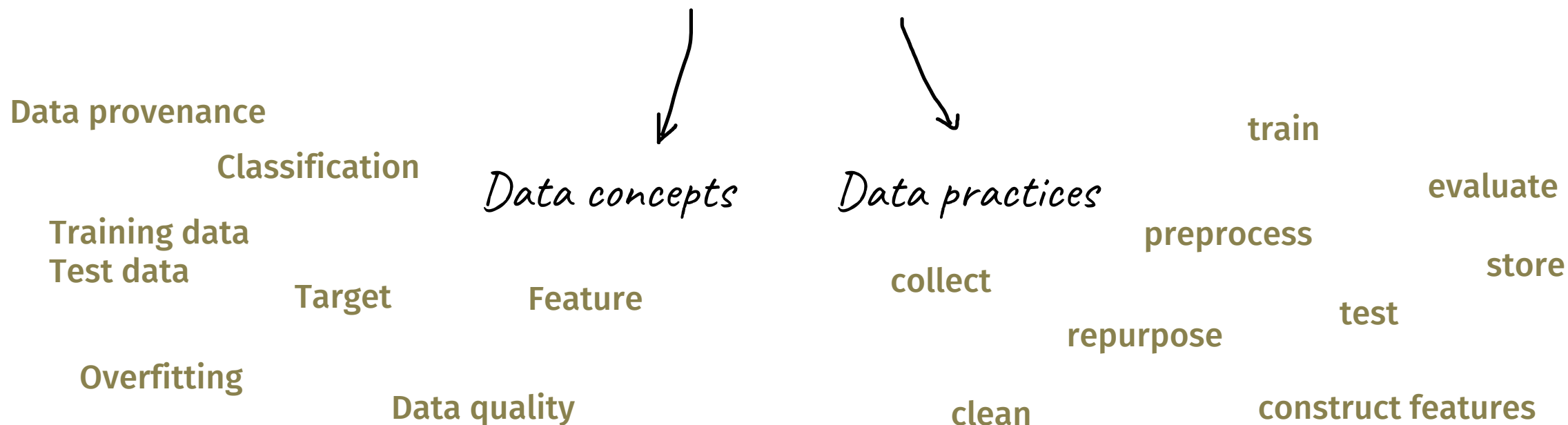
Zha et al. 2023,
Jakubik et al. 2024
Ali et al. 2023

In order to understand, evaluate, use responsibly,
and design such systems, an
understanding of data is necessary.

Motivation

Zha et al. 2023,
Jakubik et al. 2024
Ali et al. 2023

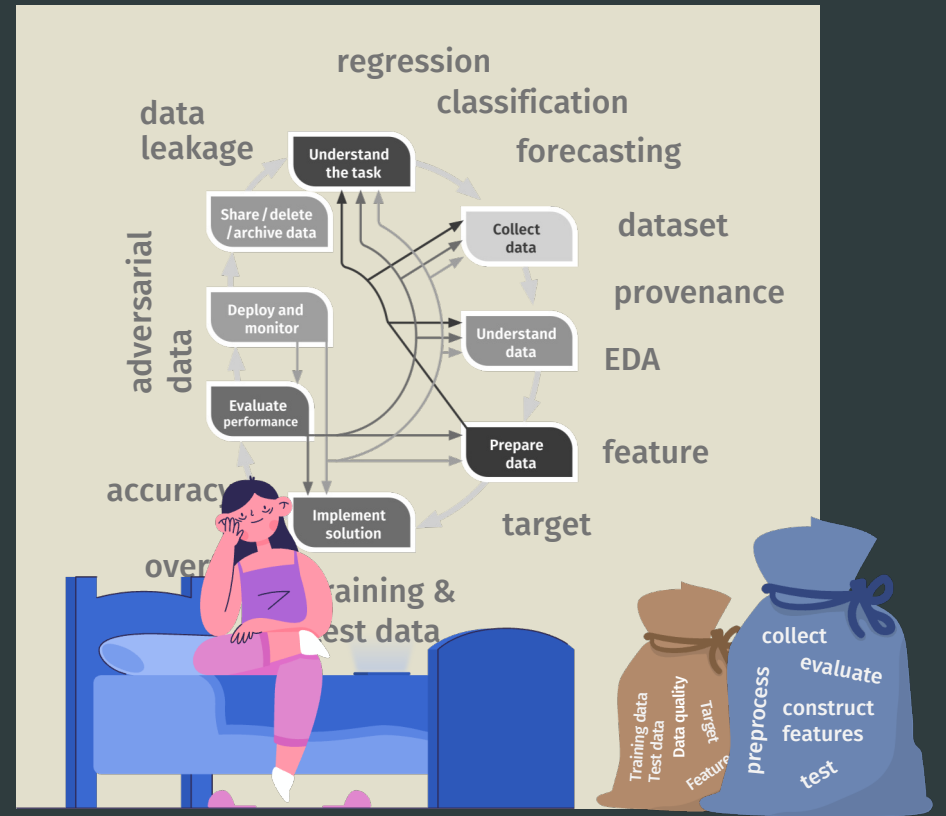
In order to understand, evaluate, use responsibly,
and design such systems, an
understanding of data is necessary.



Research question

What are the **key data-related concepts and practices** in context of AI that school students need to know?

Data concepts & practices



Methodology

4 Steps for identification of concepts & practices

corroborate

align

analyse

conceptualise

Methodology

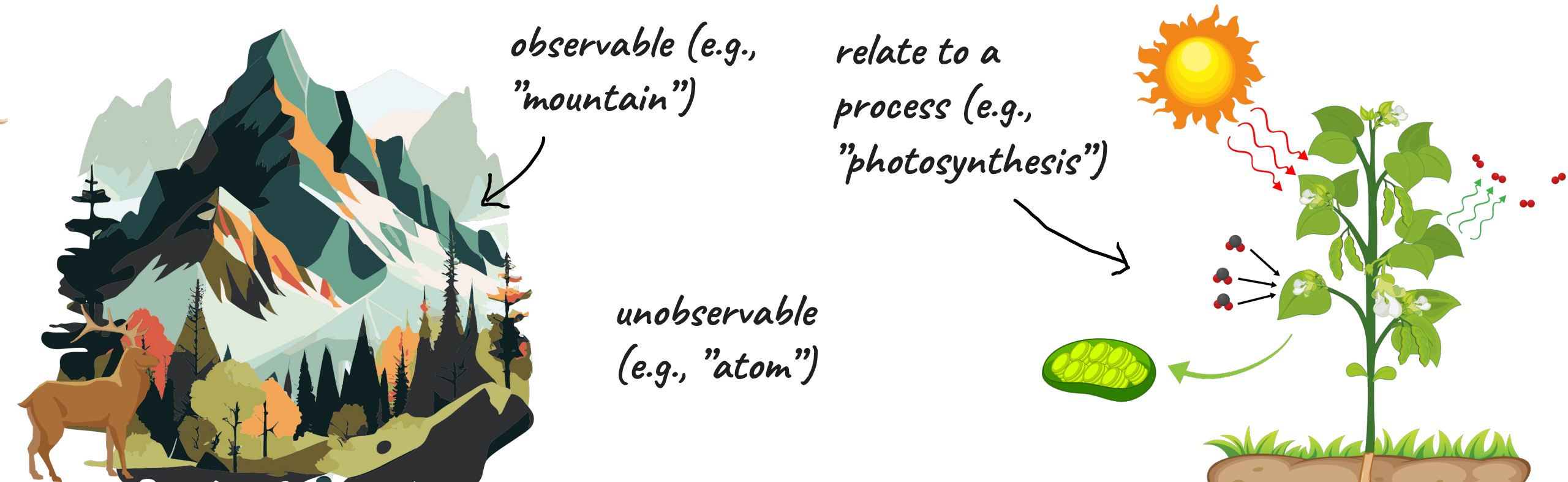
4 Steps for identification of concepts & practices

conceptualise

Concepts in sciences

Kampourakis (2018)

Concepts are described as **systematic mental representations of the real world.**

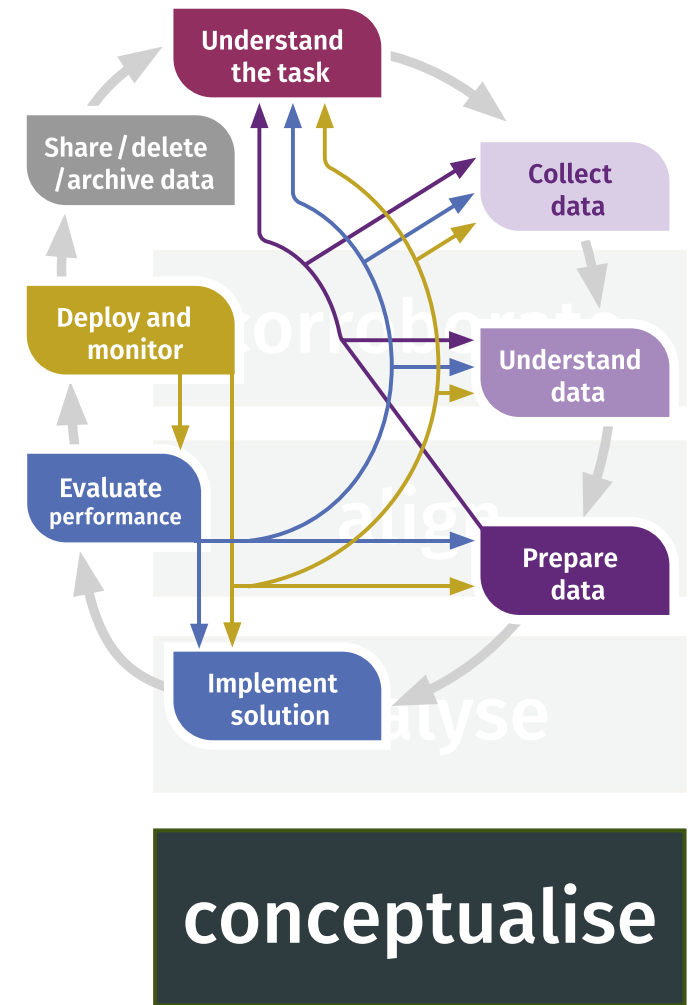


Methodology

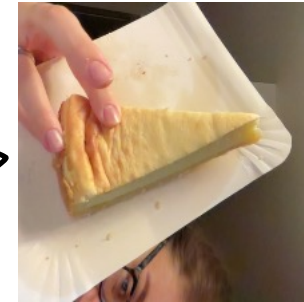
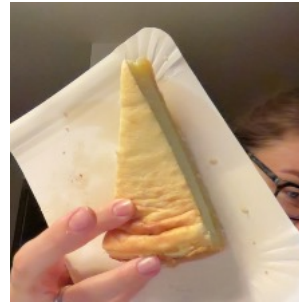
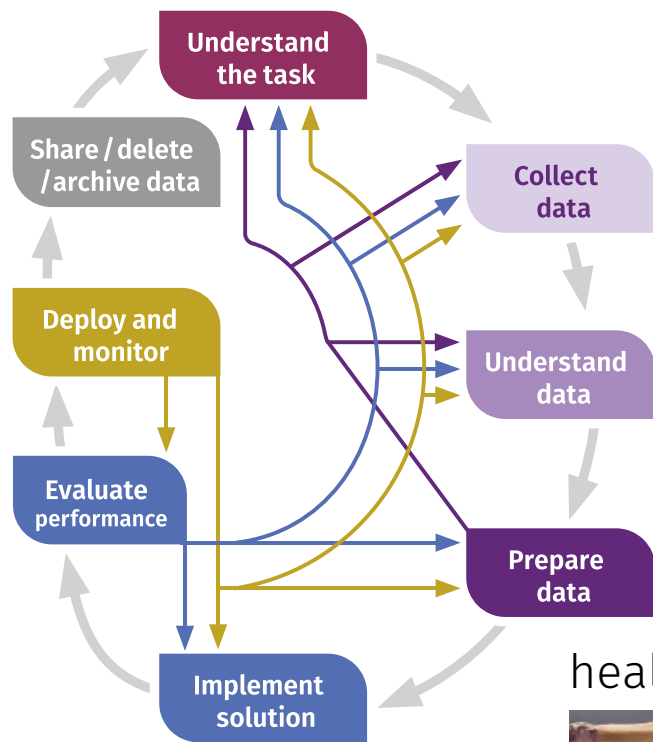
4 Steps for identification of concepts

Term used to describe a specific **process**, an **observable**, or **unobservable entity related to data** in one of the stages of the data-driven lifecycle.

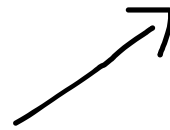
Olari & Romeike (2024)



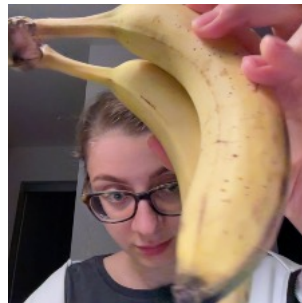
Unobservable (e.g., data modality)



Relate to a process
(e.g., data augmentation)



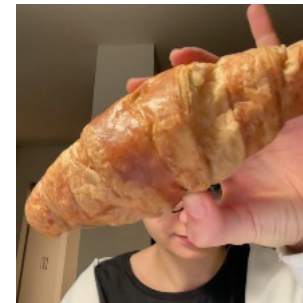
healthy_1.png



healthy_2.png



unhealthy_1.png



...

Observable (e.g.,
training data)



corroborate

align

analyse

conceptualise

Practices in sciences

Costa & Broietti (2023)

Practices are actions performed by scientists to build up knowledge.

Planning and carrying out investigations.



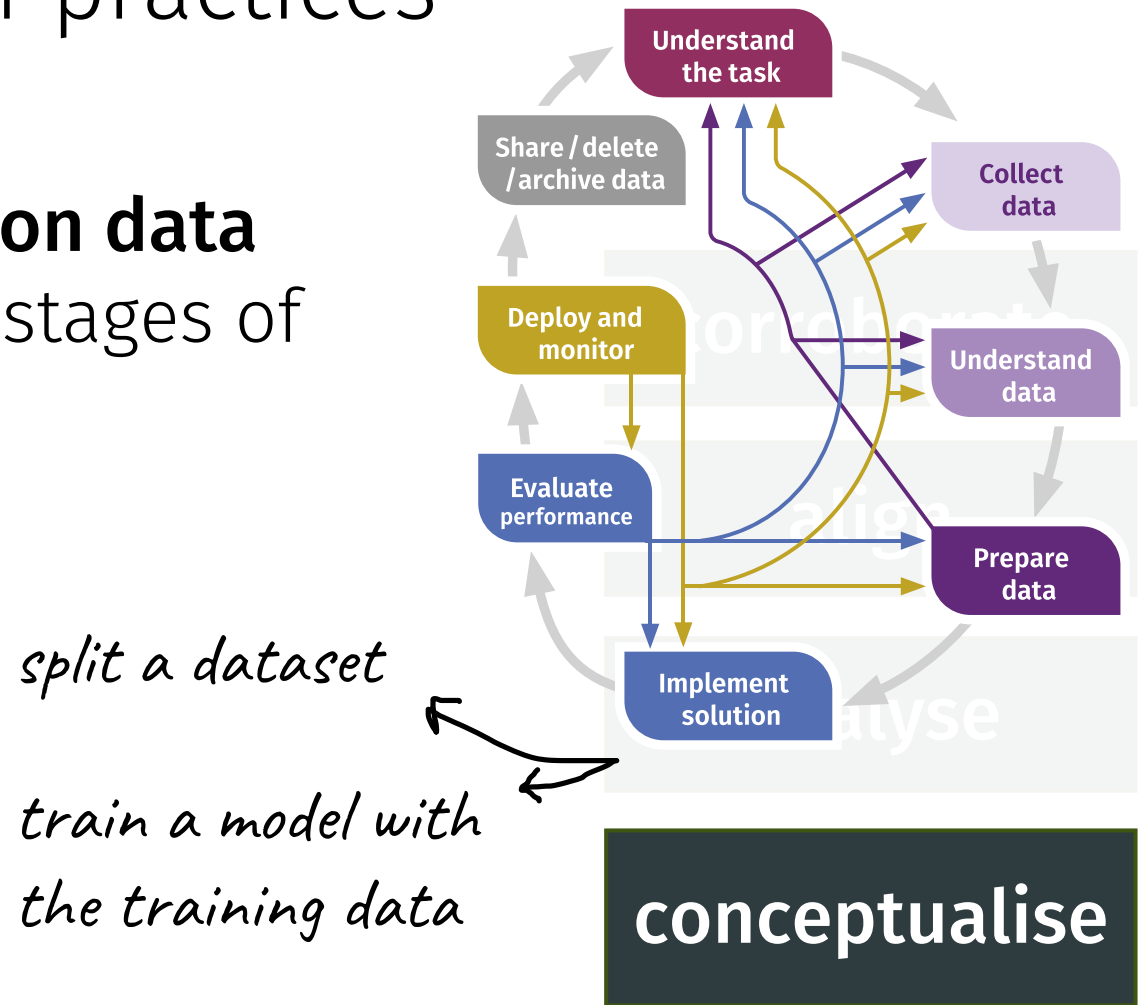
Developing and using models.

Methodology

4 Steps for identification of practices

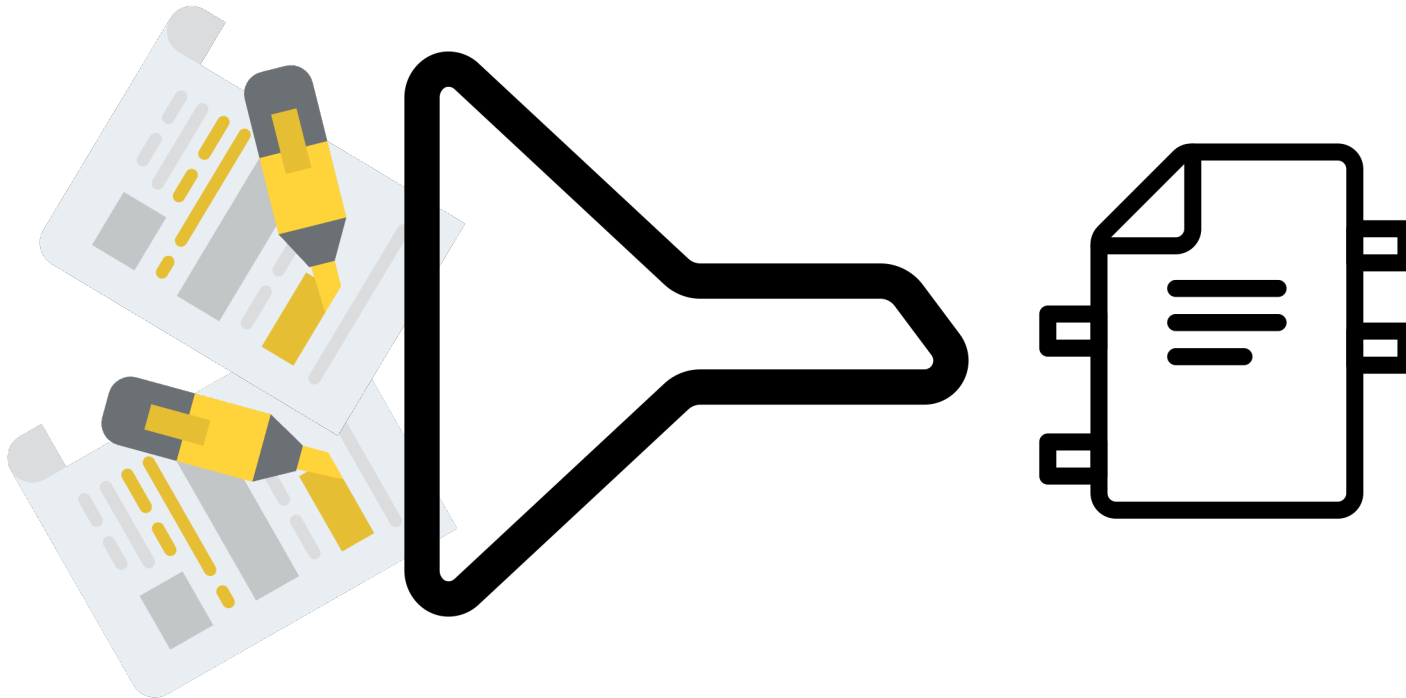
A practice is an **action performed on data** or with data in mind in one of the stages of the data-driven lifecycle.

Olari & Romeike (2024)



Methodology

4 Steps for identification of concepts & practices



corroborate

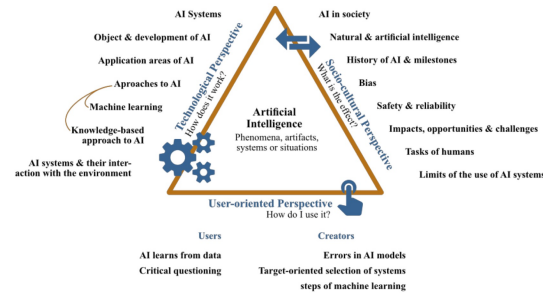
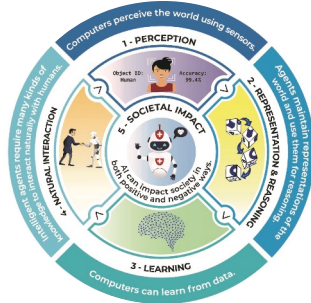
align

analyse

conceptualise

Methodology

4 Steps for identification of concepts & practices

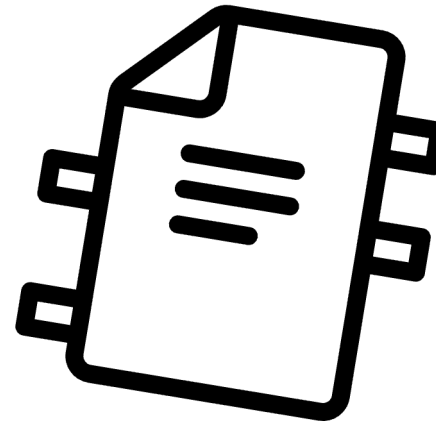


corroborate

align

analyse

conceptualise



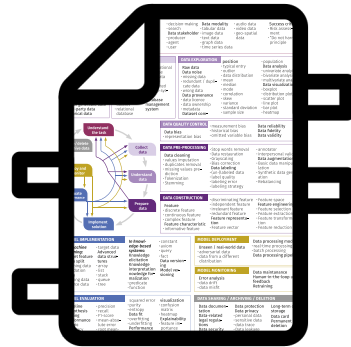
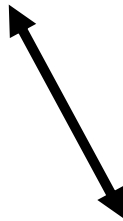
Methodology

4 Steps for identification of concepts & practices

Domain
Experts



Researchers



CS Teachers



corroborate

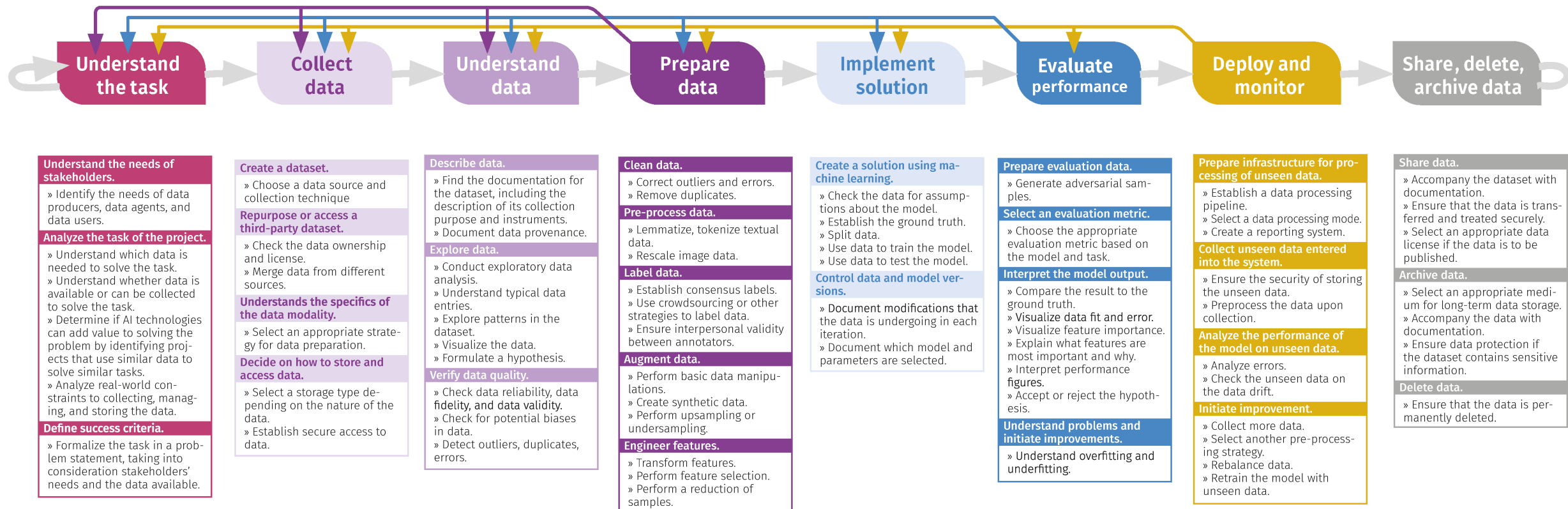
align

analyse

conceptualise

Results

An 8-stages process model

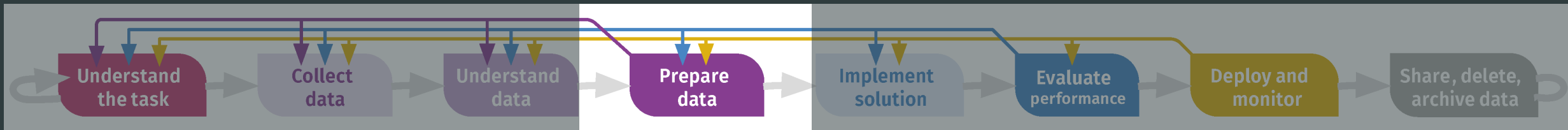


28 key data-related practices

>> 69 examples of subpractices for use in CS education

Results

An 8-stages process model



Clean data.

- » Correct outliers and errors.
- » Remove duplicates.

Pre-process data.

- » Lemmatize, tokenize textual data.
- » Rescale image data.

Label data.

- » Establish consensus labels.
- » Use crowdsourcing or other strategies to label data.
- » Ensure interpersonal validity between annotators.

Augment data.

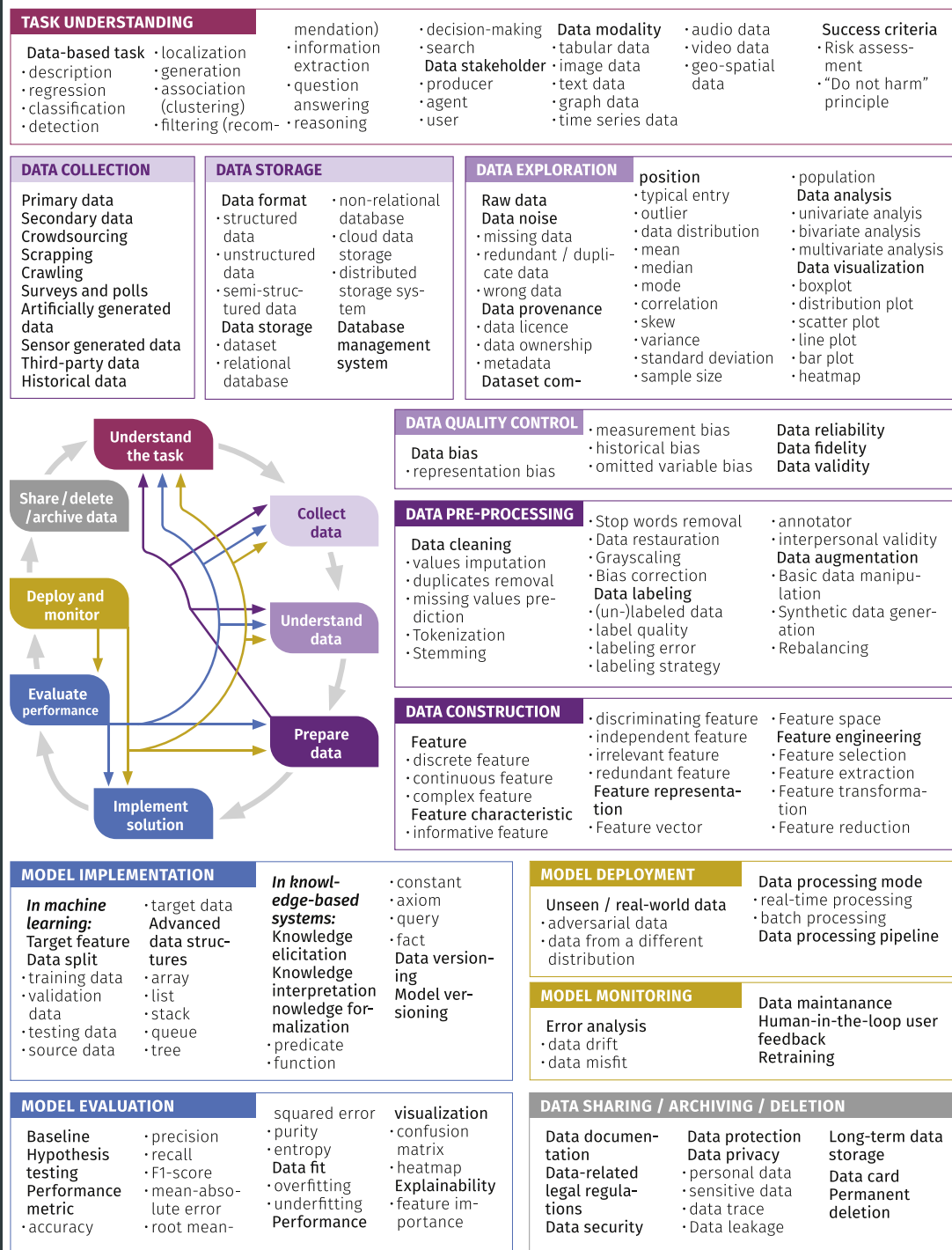
- » Perform basic data manipu-

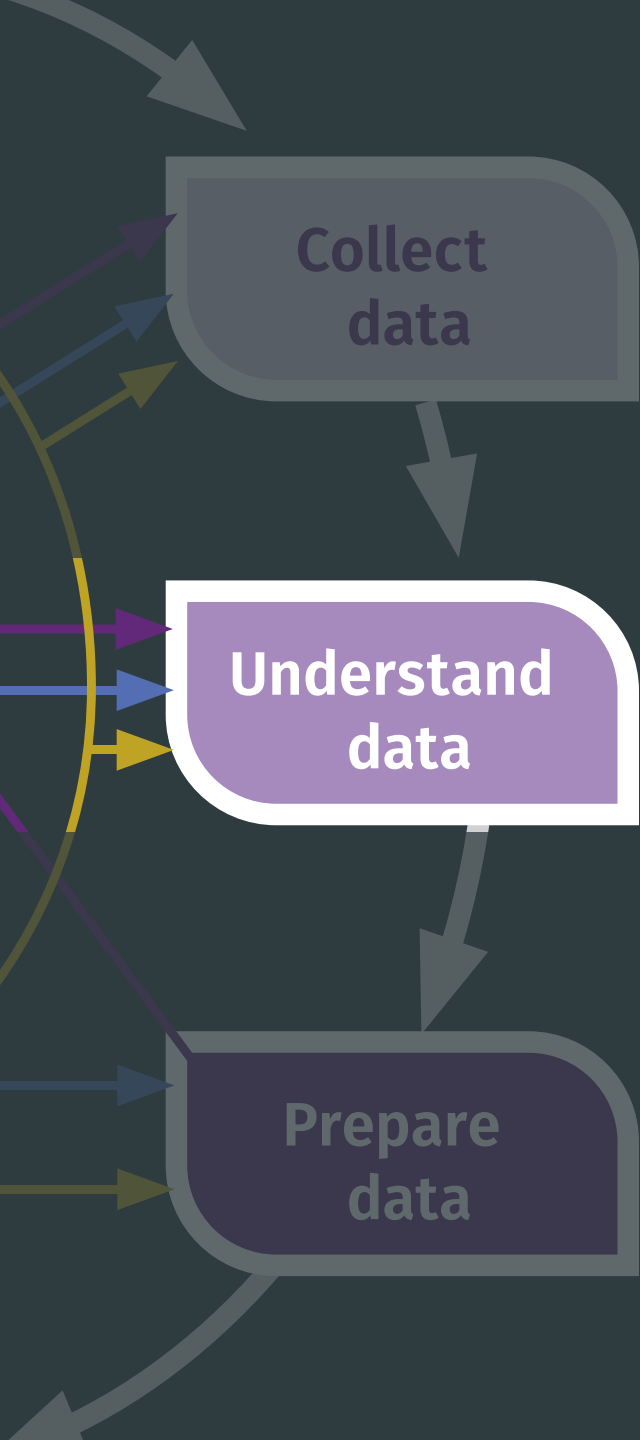
↙ *key data-related practice*

← *examples of subpractices*

Results

A collection of 133 data-related concepts





DATA EXPLORATION

Raw data

Data noise

- missing data
- redundant / duplicate data
- wrong data

Data provenance

- data licence
- data ownership
- metadata

Dataset com-

position

- typical entry
- outlier

data distribution

- mean
- median
- mode
- correlation
- skew
- variance
- standard deviation
- sample size

population

Data analysis

- univariate analysis
- bivariate analysis
- multivariate analysis

Data visualization

- boxplot
- distribution plot
- scatter plot
- line plot
- bar plot
- heatmap

DATA QUALITY CONTROL

Data bias

- representation bias

- measurement bias
- historical bias
- omitted variable bias

Data reliability

Data fidelity

Data validity

How can I use it when specifying what to teach?

A collection of
133 data-related
concepts

+

An **8-stages**
process model

... **describe** how to
design an AI system.

OECD (2025)



Go backwards and use the collection as orientation.

... **describe** how to design an AI system.

OECD (2025)

What terms should my students be able to operate with?

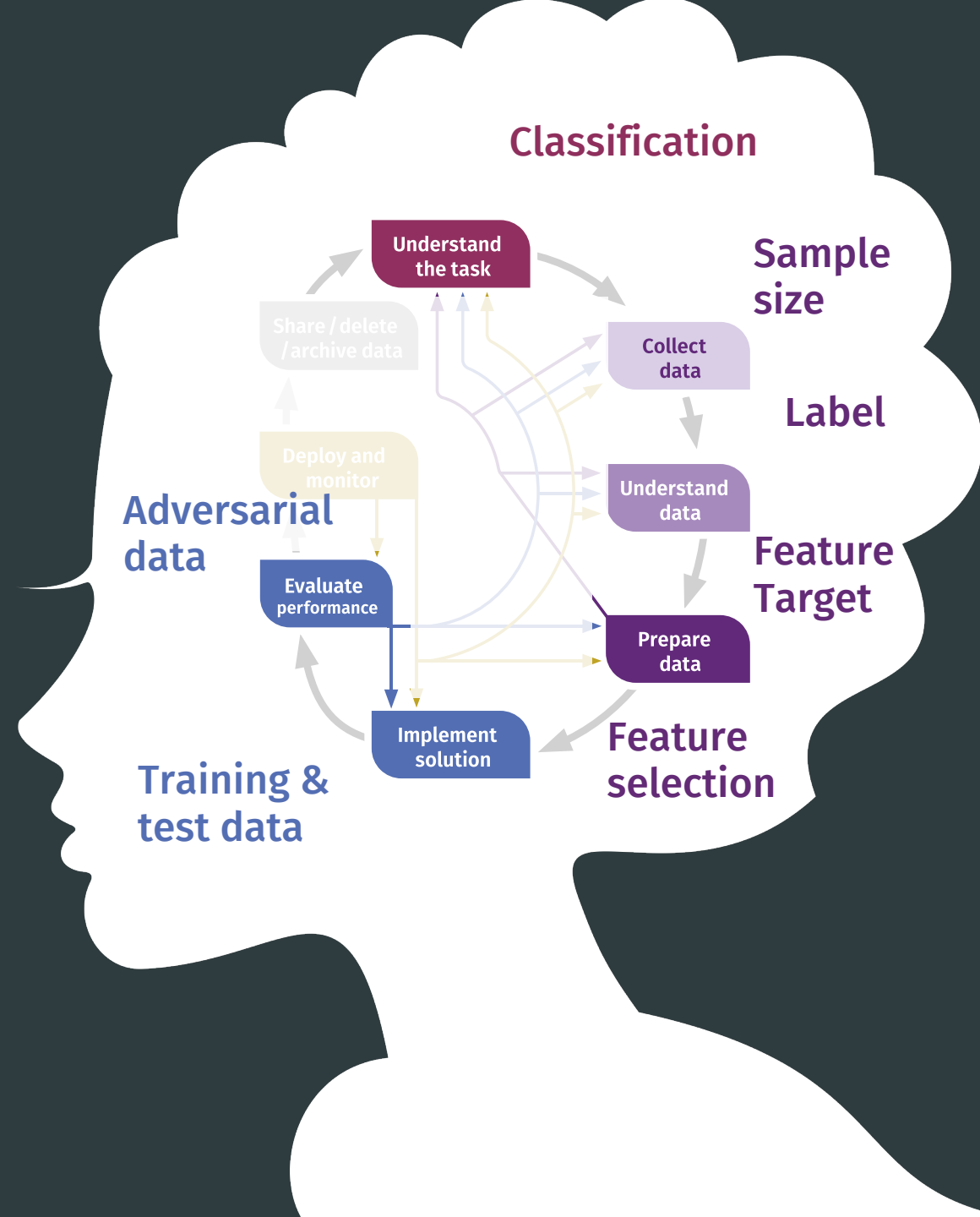
TASK UNDERSTANDING

Data-based task	• localization
• description	• generation
• regression	• association
• classification	(clustering)
• detection	• filtering (recom-

Go backwards and use the collection as orientation.

... **describe** how to design an AI system.

OECD (2025)



Go backwards and use the model as orientation.

... **describe** how to design an AI system.

OECD (2025)

Which practices should my students understand?

Clean data.

- » Correct outliers and errors.
- » Remove duplicates.

Pre-process data.

- » Lemmatize, tokenize textual data.
- » Rescale image data.

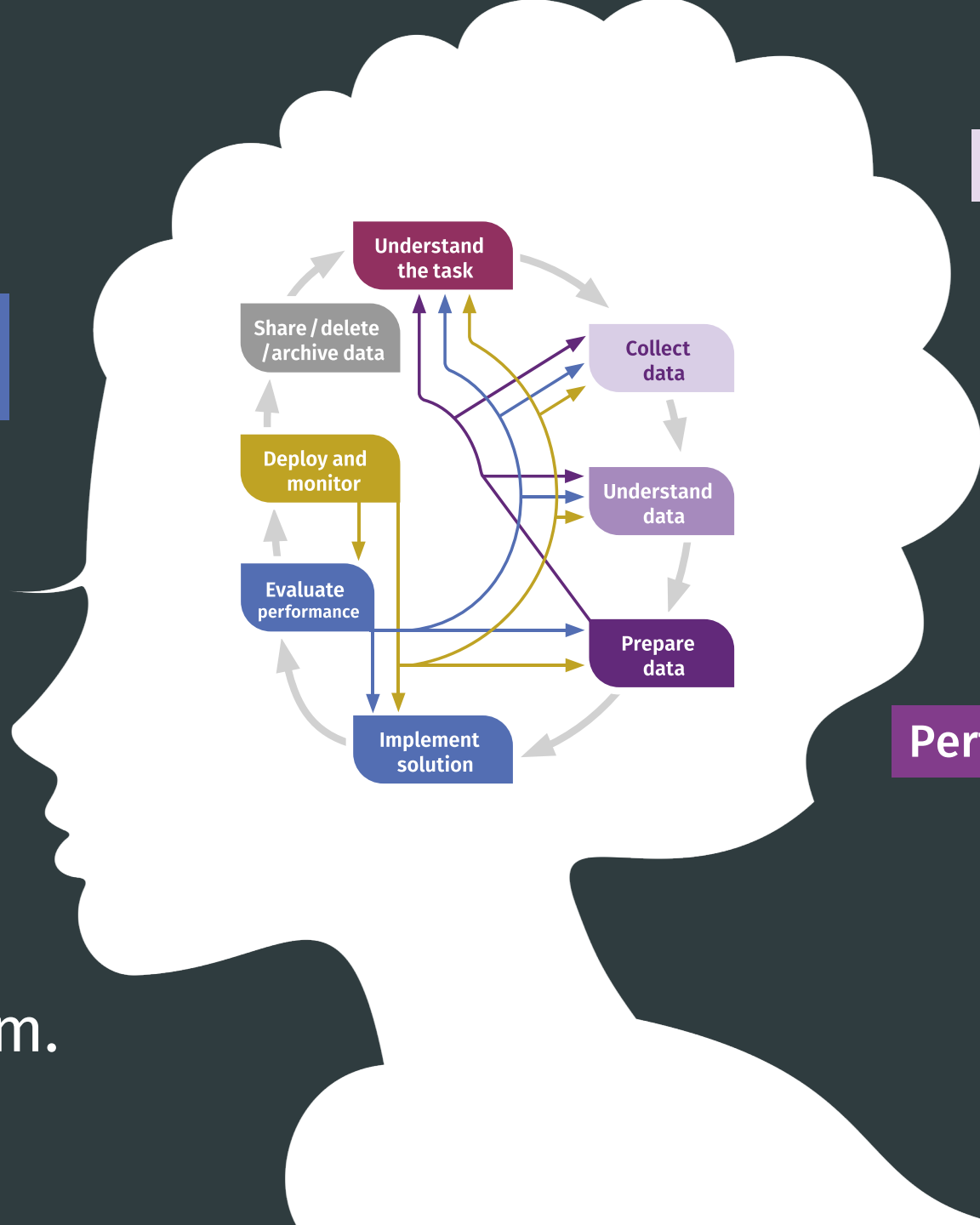
Label data.

Generate adversarial data to test robustness.

Use training data to train the model.

... **describe** how to design an AI system.

OECD (2025)



Create a dataset.

Verify data quality.

Label data.

Perform feature selection.

The student should be able to operate with the following vocabulary:

Classification

Sample size

Label

**Feature
selection**

**Feature
Target**

Adversarial data

Training & testing data

The student should demonstrate understanding of the following practices:

Create a dataset.

Verify data quality.

Label data.

Perform feature selection.

**Generate adversarial
data to test robustness.**

**Use training data to
train the model.**

**... describe how to
design an AI system.**

OECD (2025)

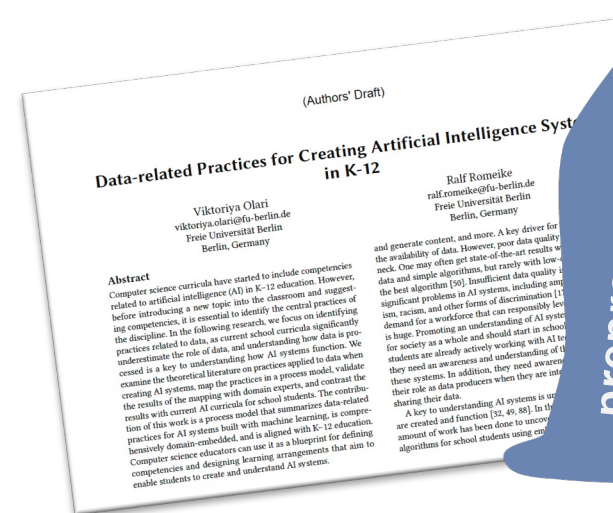
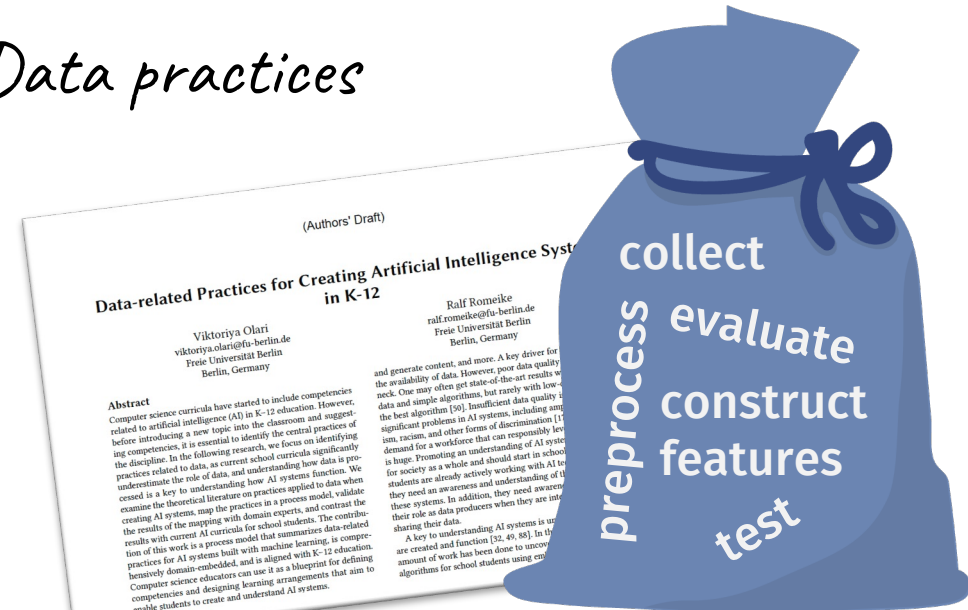
To understand such systems, you have to understand data.

Data concepts

Data practices



<https://doi.org/10.1016/j.caeo.2024.100196>



<https://doi.org/10.1145/3677619.367811>

Take away so far

A collection of concepts



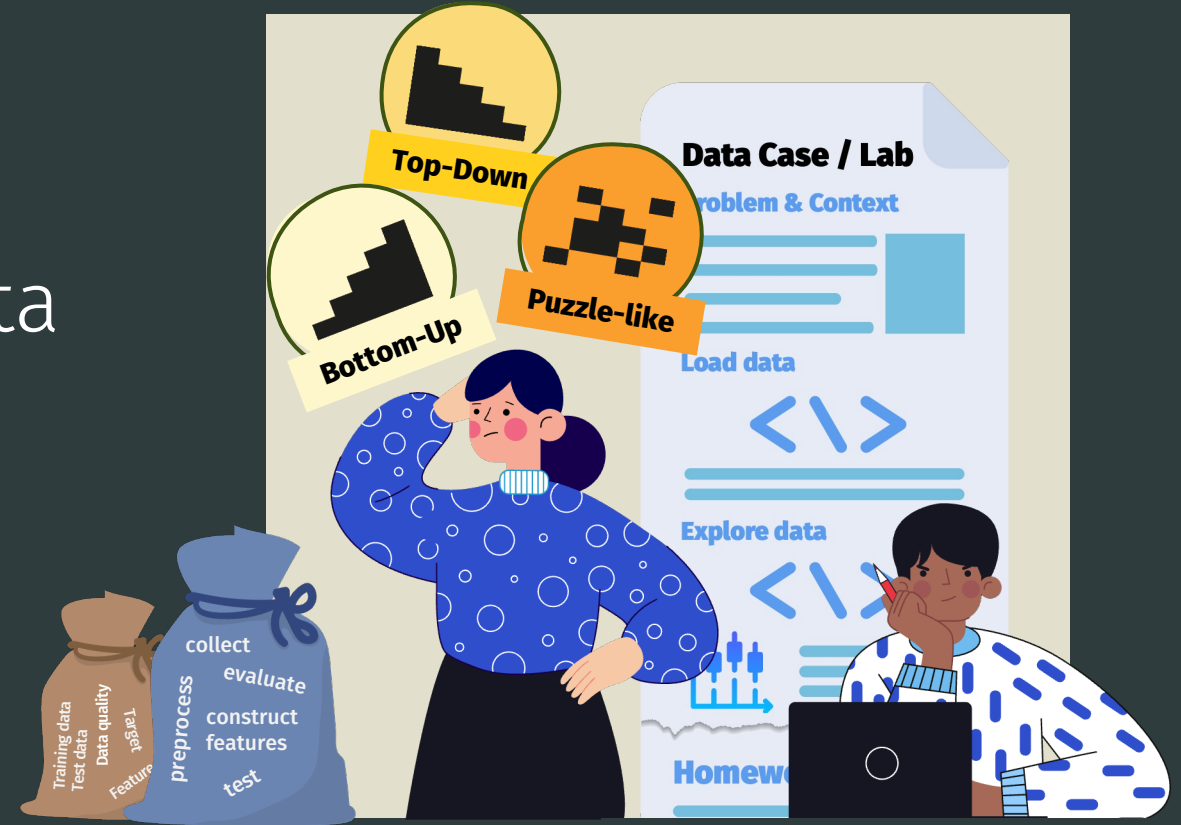
A process model

Can be used as **a foundation for a common language when teaching about data-driven systems.**

Can be used by teachers / curriculum developers **as a guide when specifying competencies.**

The Data Case Study

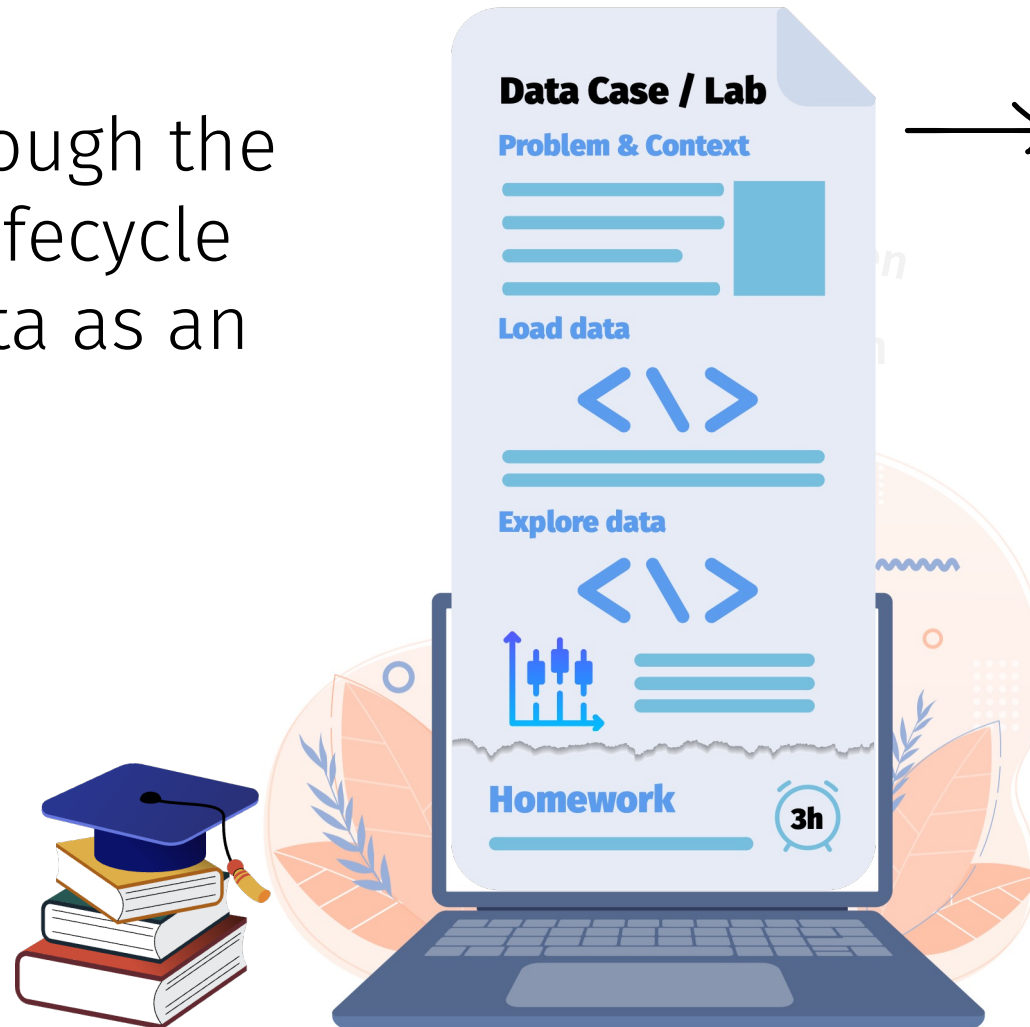
How can we support computer science teachers in addressing these **additional aspects** of data in the classroom?



The data case study

An established method in data science education

Guidance through the data-driven lifecycle using real data as an example



→ ***How will particulate matter pollution develop in Berlin over the next 10 years?***

Nolan & Speed (2001)
Chkoniya (2021)
Wright et al. (2024)

The data case study

Learning theory perspective

Situated learning

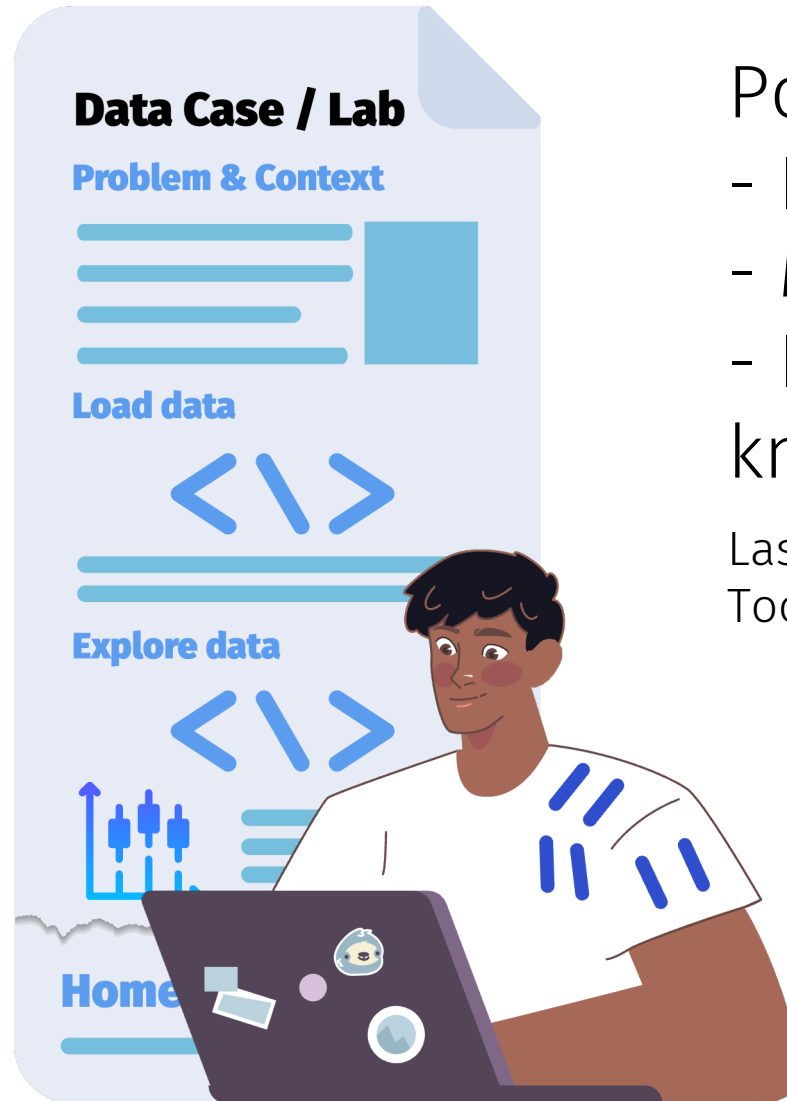
Hoffer (2020)

Constructivist
learning

Gill (2011), Hoffer (2020)

Active learning

Herreid (2011)



Positive effect on

- Interest
- Motivation
- Increase in knowledge and skills

Lasser (2021)

Toogood (2023)

The data case study

School-specific challenges in the transfer process

Lack of programming skills

Lack of domain knowledge

The time for preparing the data cases is limited.



Research Questions

What challenges need to be considered in order to effectively use the data case study as a teaching and learning method in computer science lessons on the topic of “AI”?

Research Questions

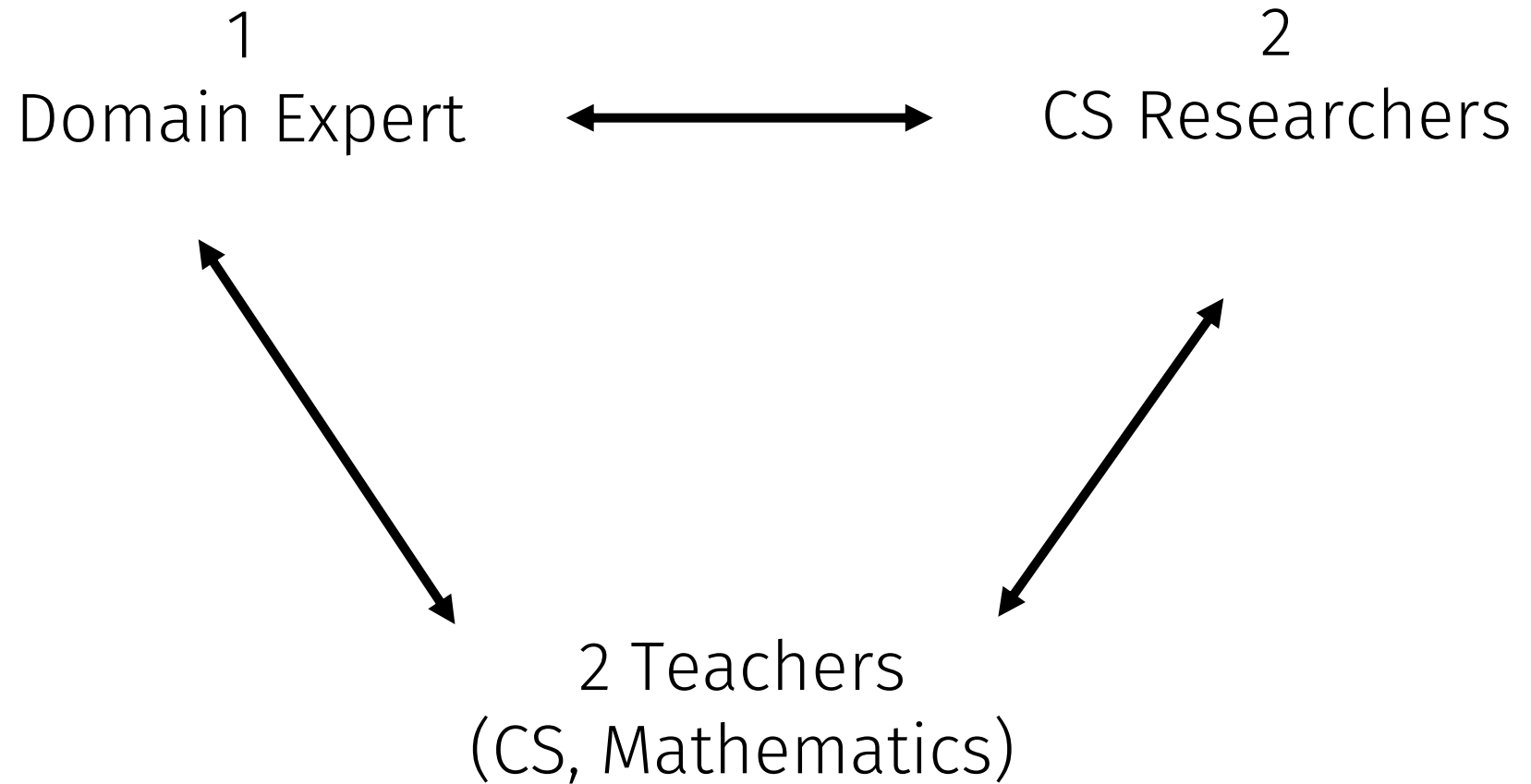
What challenges need to be considered in order to effectively use the data case study as a teaching and learning method in computer science lessons on the topic of “AI”?

How must the data case study be adapted as a teaching and learning method in order to overcome the challenges identified and meet the objectives of computer science education on the topic of “AI”?

Methodology

Methodology

Expert Team



Methodology

Design-based research

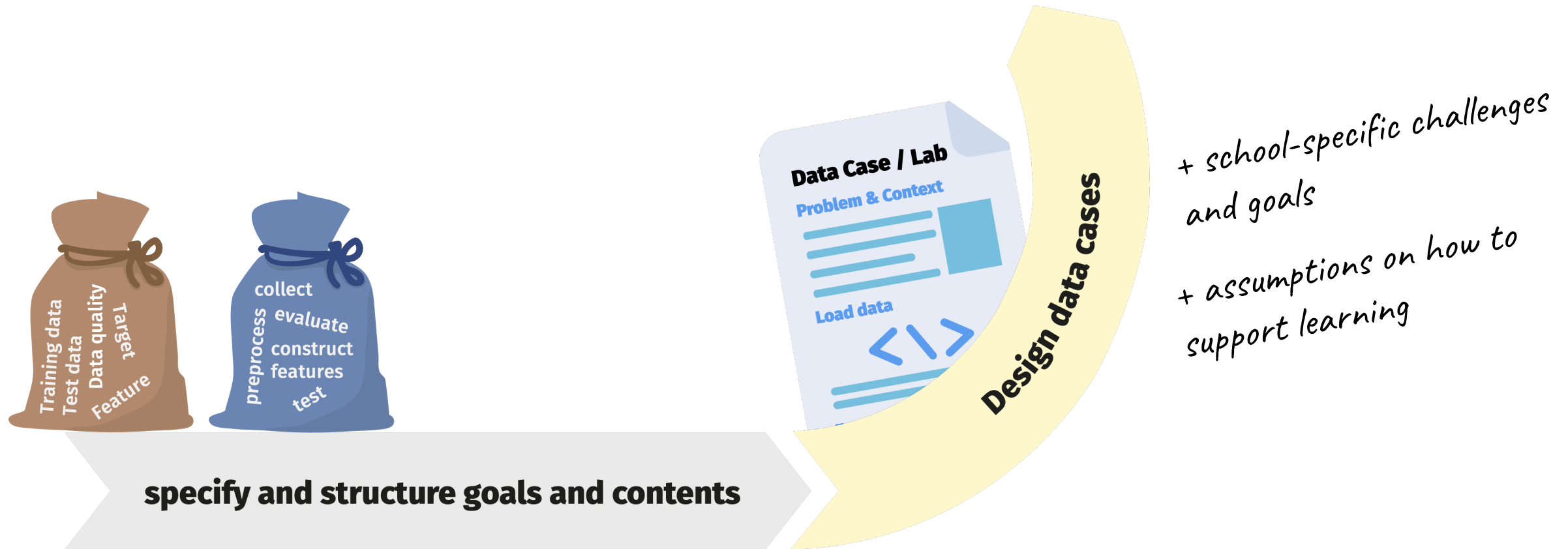
Design-based research

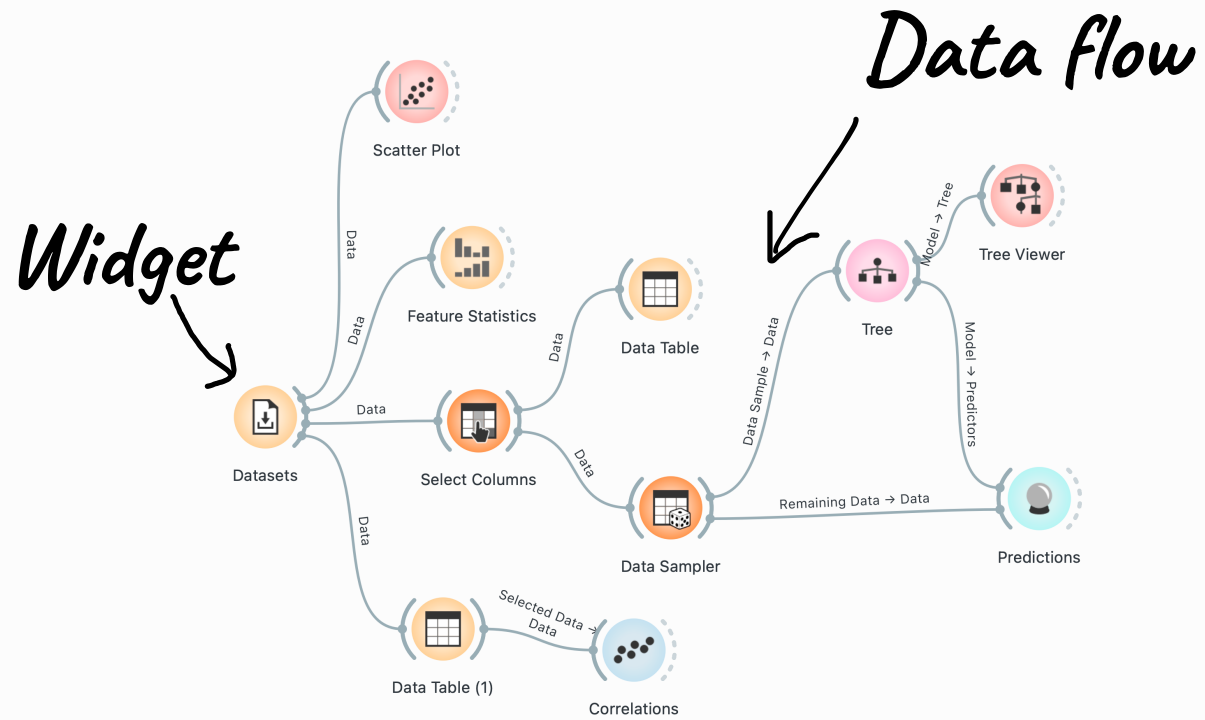
(Prediger 2012)



Design-based research

(Prediger 2012)





Flow-based

Design data cases

Let's also take a look to see how many monitors there are in a few cities. We can use the `filter()` function of the `dplyr` package to do so. For example, let's look at Albuquerque, New Mexico.

```
pm %>% dplyr::filter(city == "Albuquerque")
```

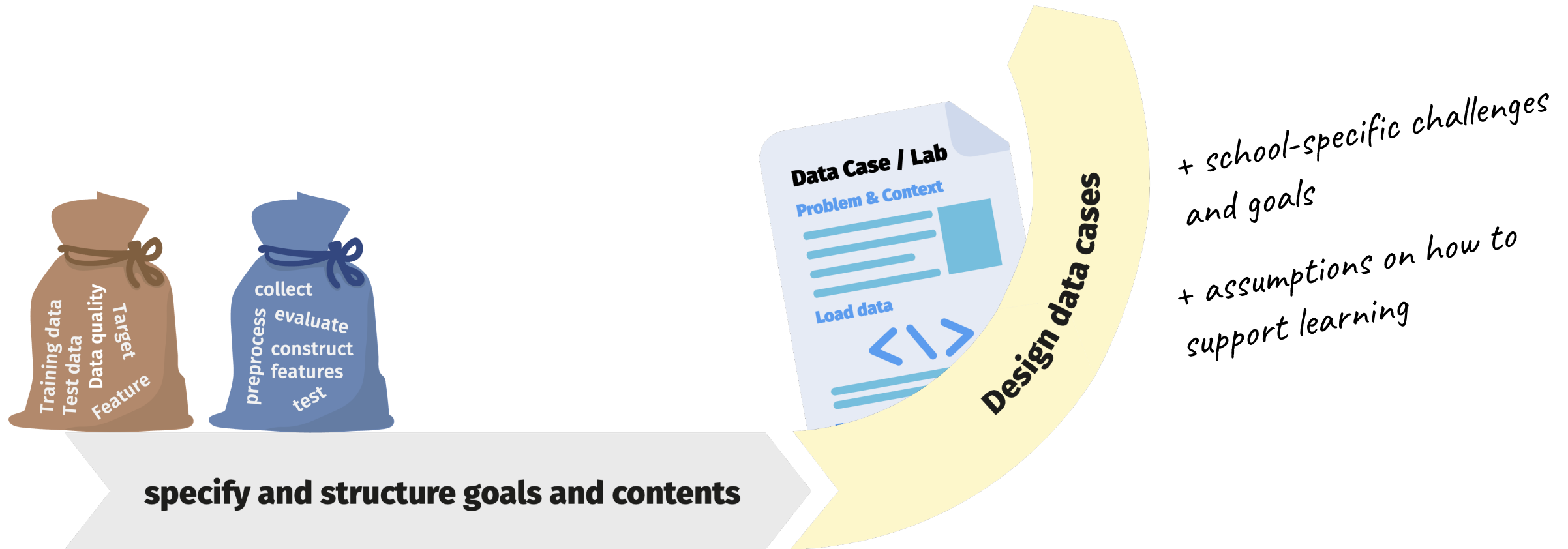
```
# A tibble: 2 × 50
  id      value fips    lat    lon state county city    CMAQ zcta  zcta_...¹ zcta_...²
<fct>   <dbl> <fct> <dbl> <dbl> <chr> <chr> <chr> <dbl> <fct> <dbl> <dbl>
1 35001...  5.98 35001  35.1 -107. New ... Berna... Albu... 10.1 87109 2.62e7 40432
2 35001...  5.91 35001  35.1 -107. New ... Berna... Albu... 10.1 87108 1.52e7 38647
# ... with 38 more variables: imp_a500 <dbl>, imp_a1000 <dbl>, imp_a5000 <dbl>,
#   imp_a10000 <dbl>, imp_a15000 <dbl>, county_area <dbl>, county_pop <dbl>,
#   log_dist_to_prisec <dbl>, log_pri_length_5000 <dbl>,
#   log_pri_length_10000 <dbl>, log_pri_length_15000 <dbl>,
#   log_pri_length_25000 <dbl>, log_prisec_length_500 <dbl>,
#   log_prisec_length_1000 <dbl>, log_prisec_length_5000 <dbl>,
#   log_prisec_length_10000 <dbl>, log_prisec_length_15000 <dbl>, ...
# i Use `colnames()` to see all variable names
```

We can see that there were only two monitors in the city of Albuquerque in 2006. Let's compare this with Baltimore.

Text-based

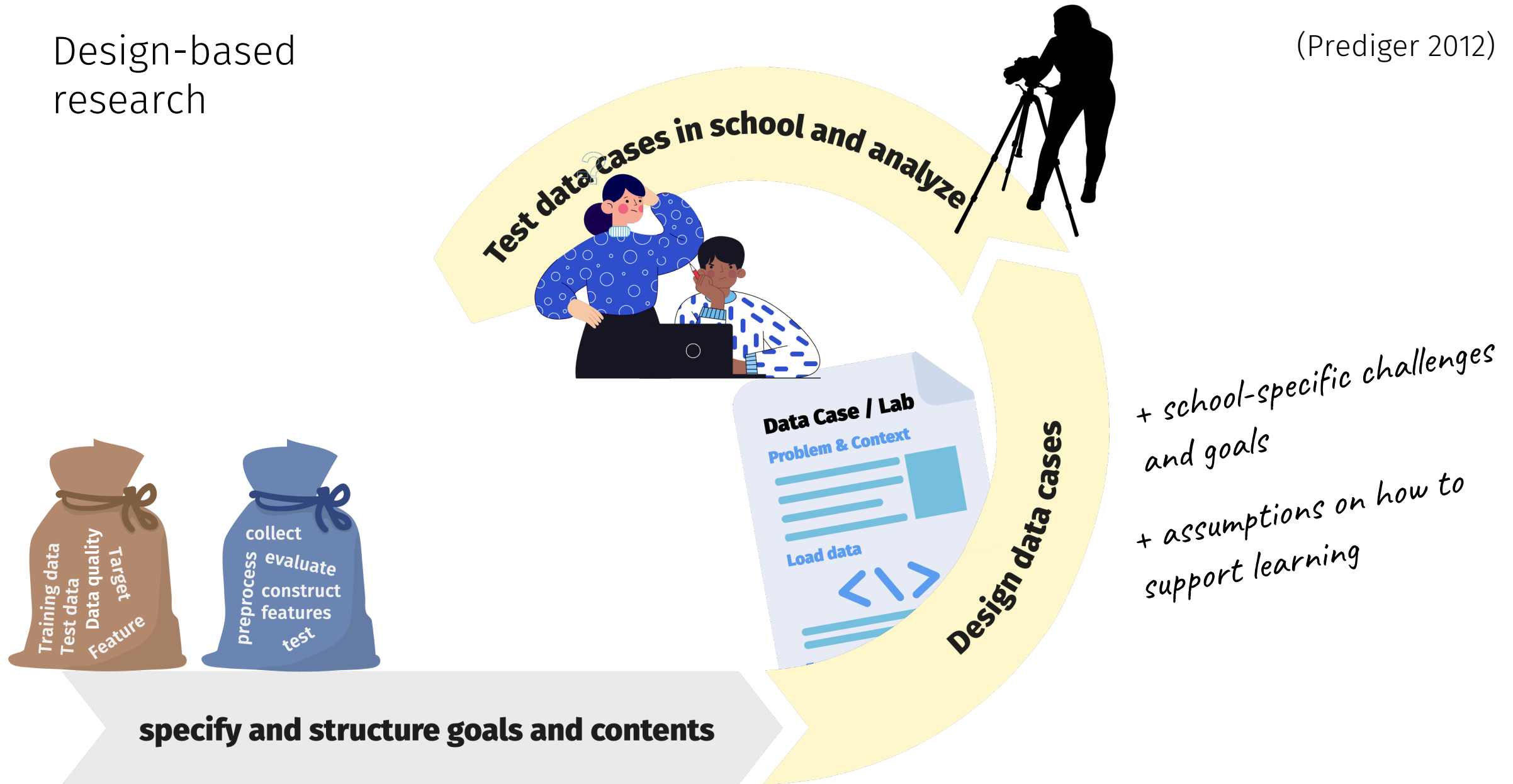
Design-based research

(Prediger 2012)



Design-based research

(Prediger 2012)



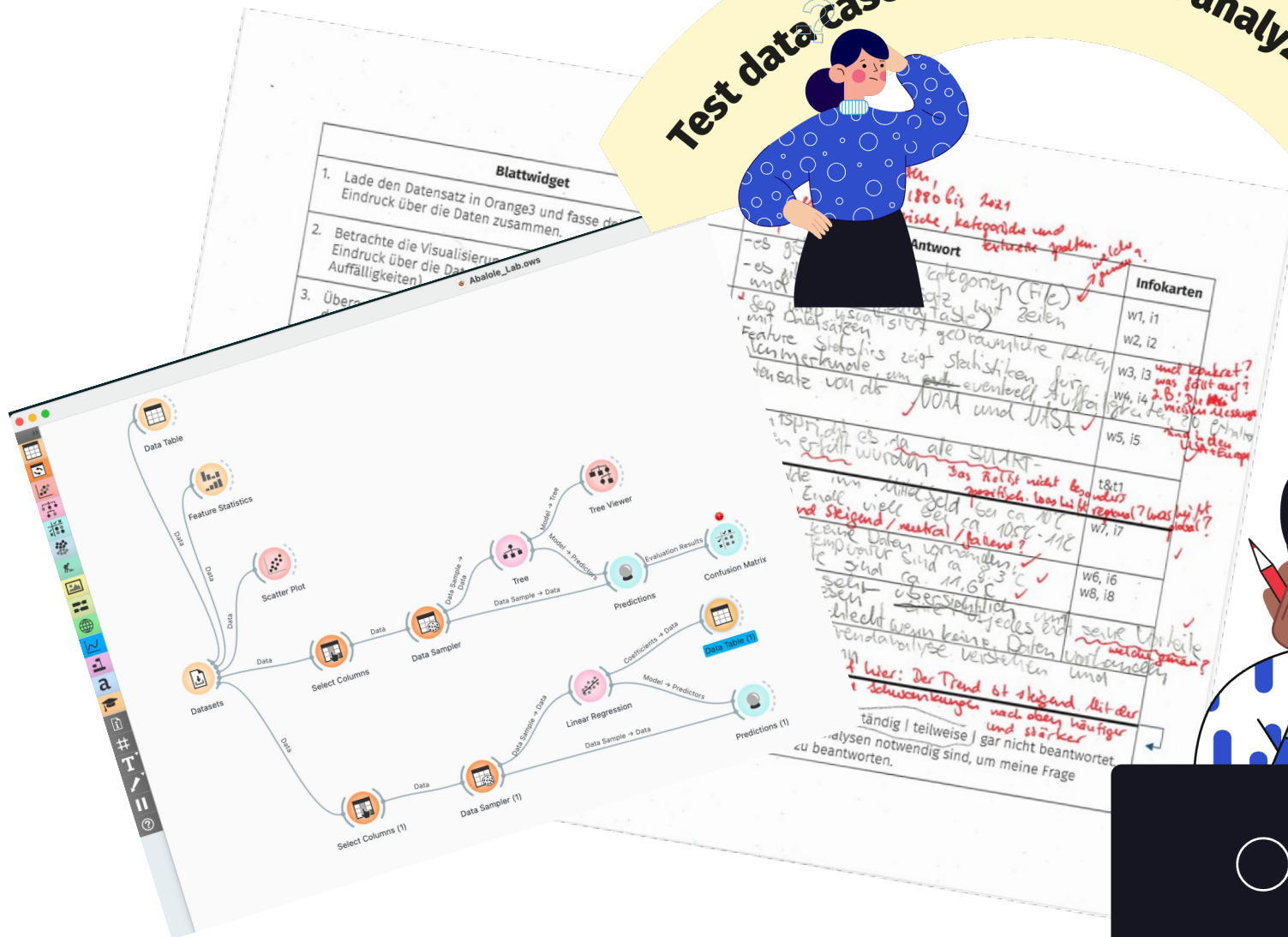
Interviews with teachers



What difficulties did students experience with the data case, and how did they overcome them?

Artifacts made by students

Test data cases in school and analyze



Die bestgelösten Aufgaben:

1. Einordnung von Tree/Linear Regression im Prozess.
2. Zuordnungsaufgabe (Ordne Tabellen dem Datenfluss zu).
3. Ankreuzen, um welches Problem es sich handelt (Regression / Klassifikation).
4. + dazugehörige Begründung.

Die schwierigsten Aufgaben waren:

1. Beschreibung der Aufgaben von Tree und Linear Regression
2. Zuordnung von Widgets zu Frage 1 und Frage 2.
3. Die meisten Orange3 Dateien hatten Fehler.

Fehler 1: SuS können nicht gut die Widgets zu Problemen zuordnen.

Unklar, wie man den Unterschied zwischen zwei kategorischen Variablen (giftig/essbar) untersuche (eigentlich mit Feature Statistics). schreibt: mit Line Chart, Scatter Plot. (Gruppe C, Teil 2, Aufgabe 1 a).

- : die Widgets für Aufgabe 1 und 2 sind identisch.
- : die Widgets für Aufgabe 1 und 2 sind identisch.

Fehler 2: Hilfematerial hilft nicht.

Aus dem Bericht: „Bei meinem Experiment habe ich durch mein Hilfsmaterial bemerkt, das ich den Tree und Tree Viewer für Aufgabe 1 benötige. Außerdem benötige ich kein Data Sampler.“
es geht doch ein Regressionsproblem

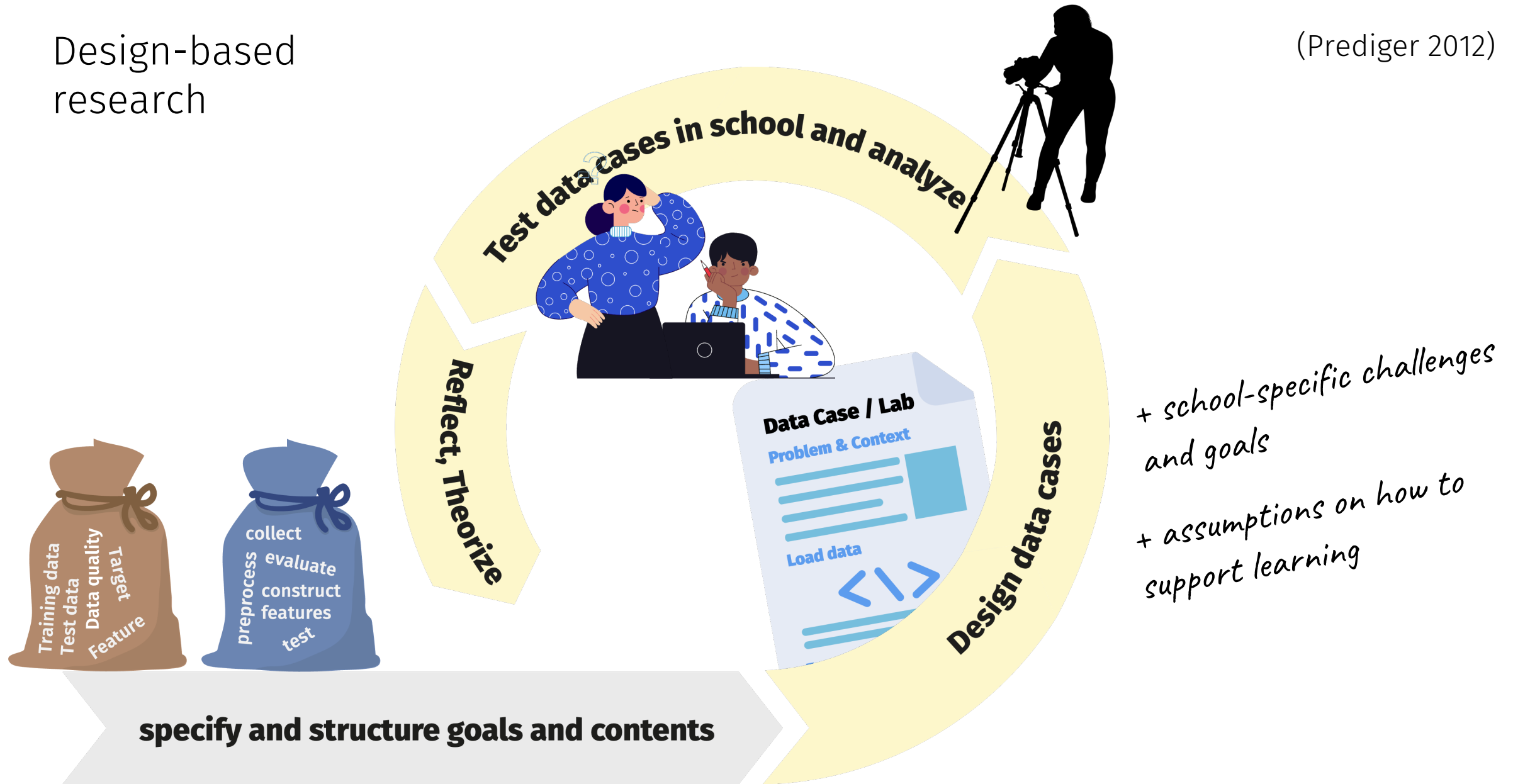
Fehler 3: Die Aufgabe nicht verstanden, Target falsch.

Zwar sind die Widgets korrekterweise miteinander verbunden, doch die Targetvariable und die Prädiktoren sind leider ganz falsch.



Design-based research

(Prediger 2012)

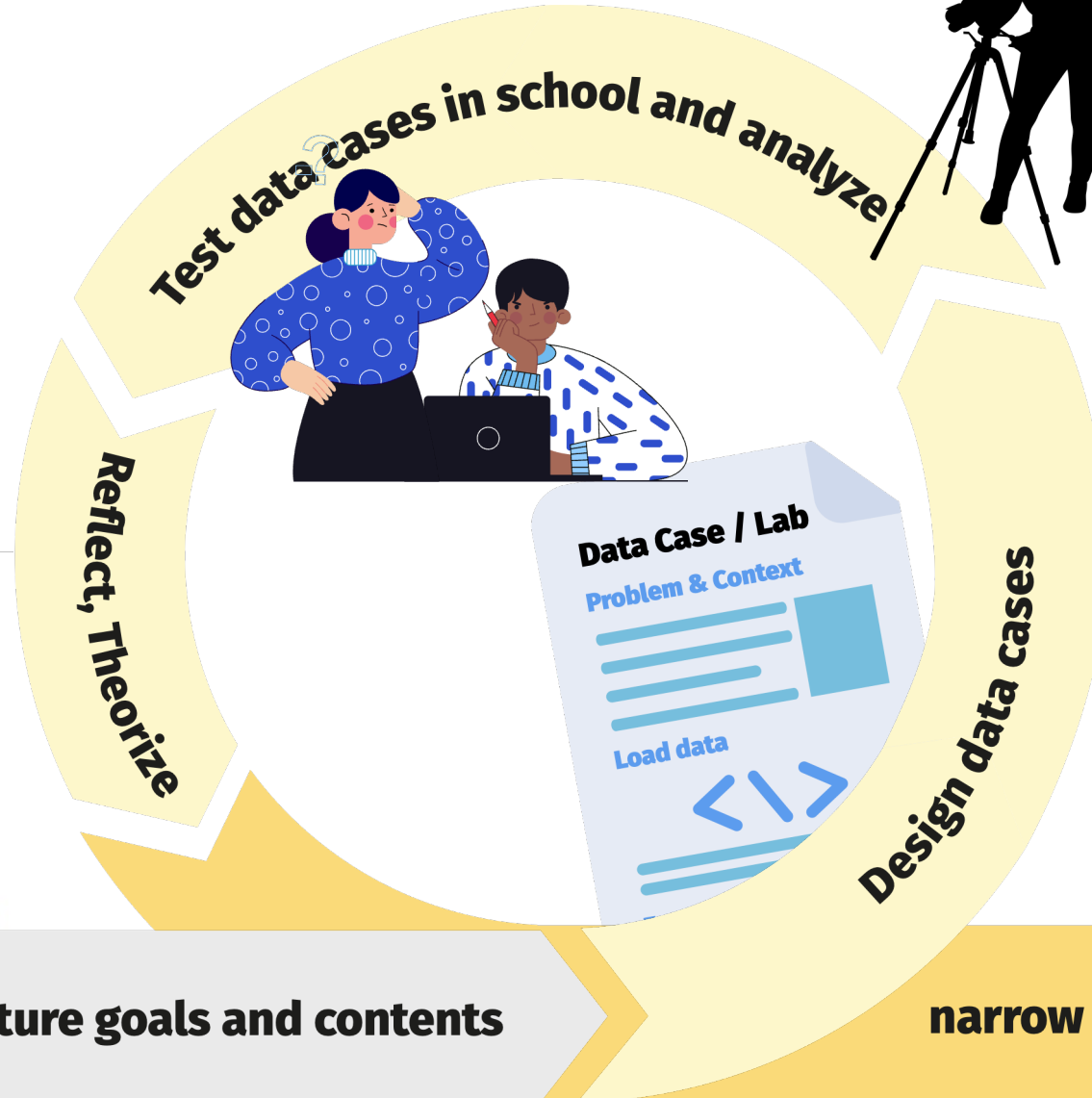


Design-based research

(Prediger 2012)

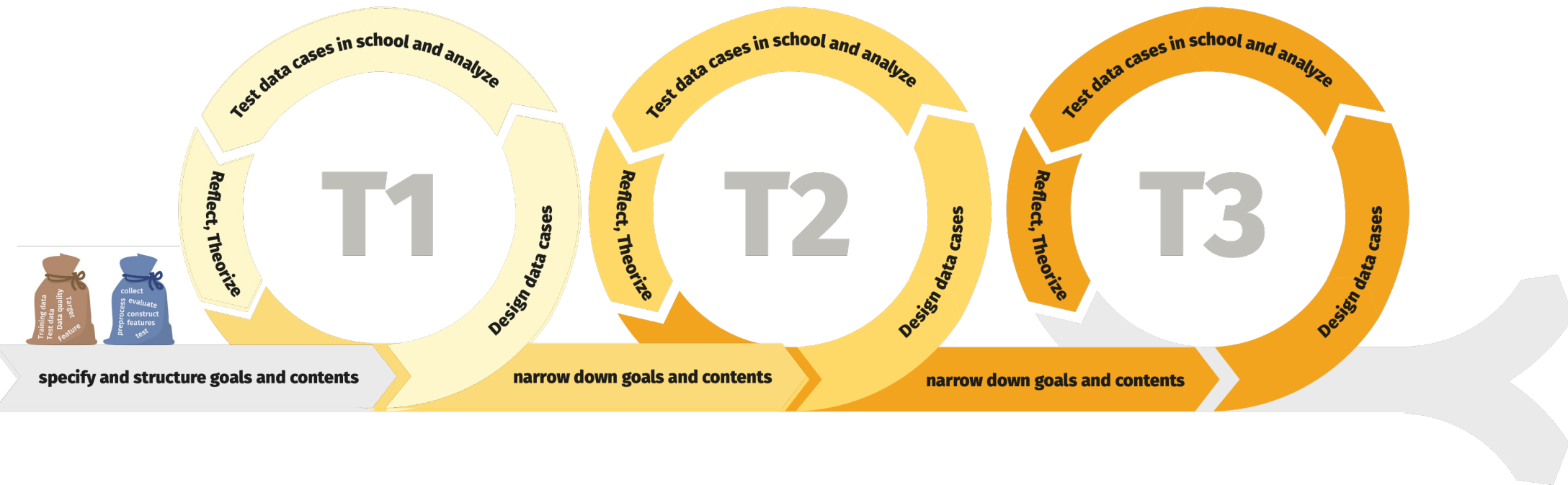


specify and structure goals and contents

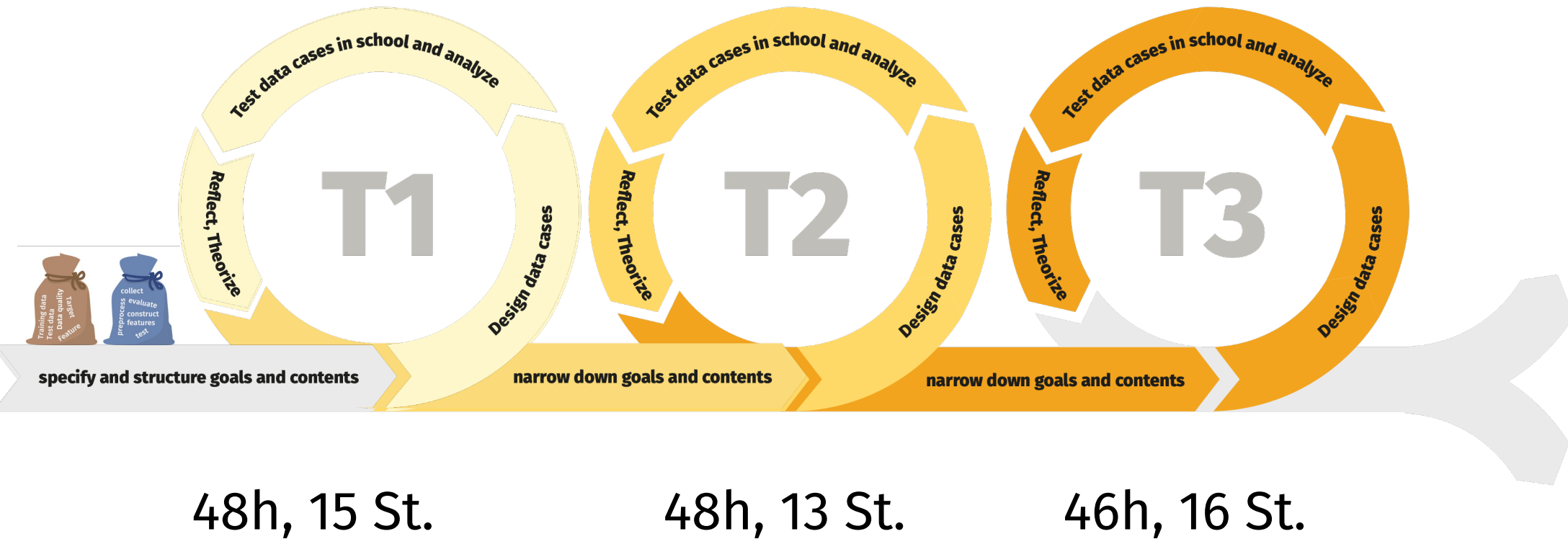


+ school-specific challenges and goals
+ assumptions on how to support learning

Design-based research



Design-based research



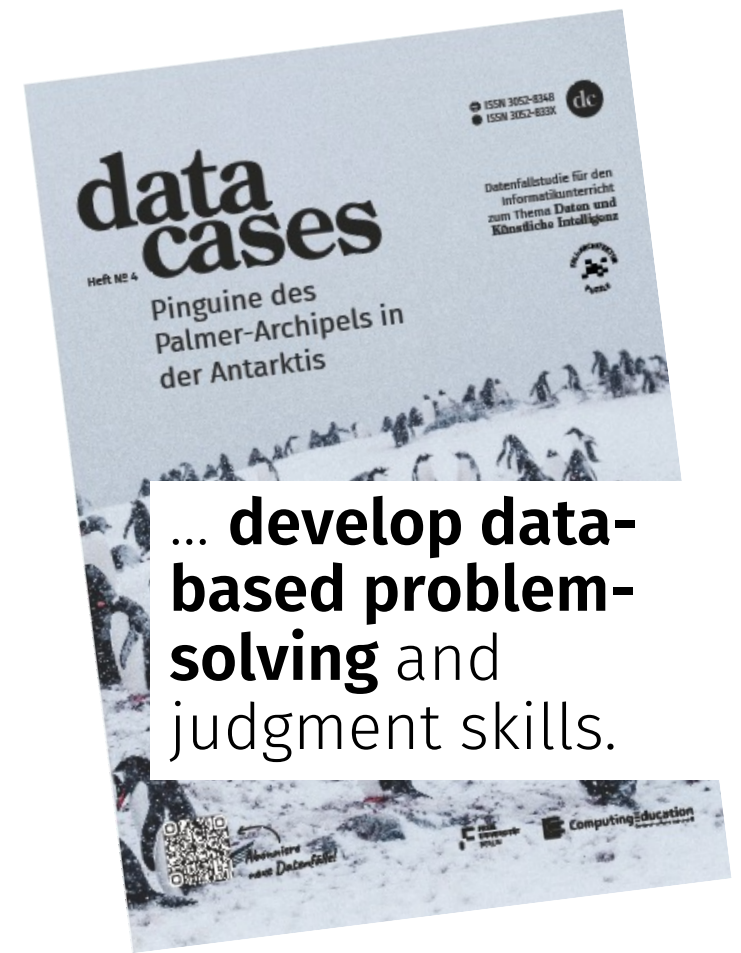
Results



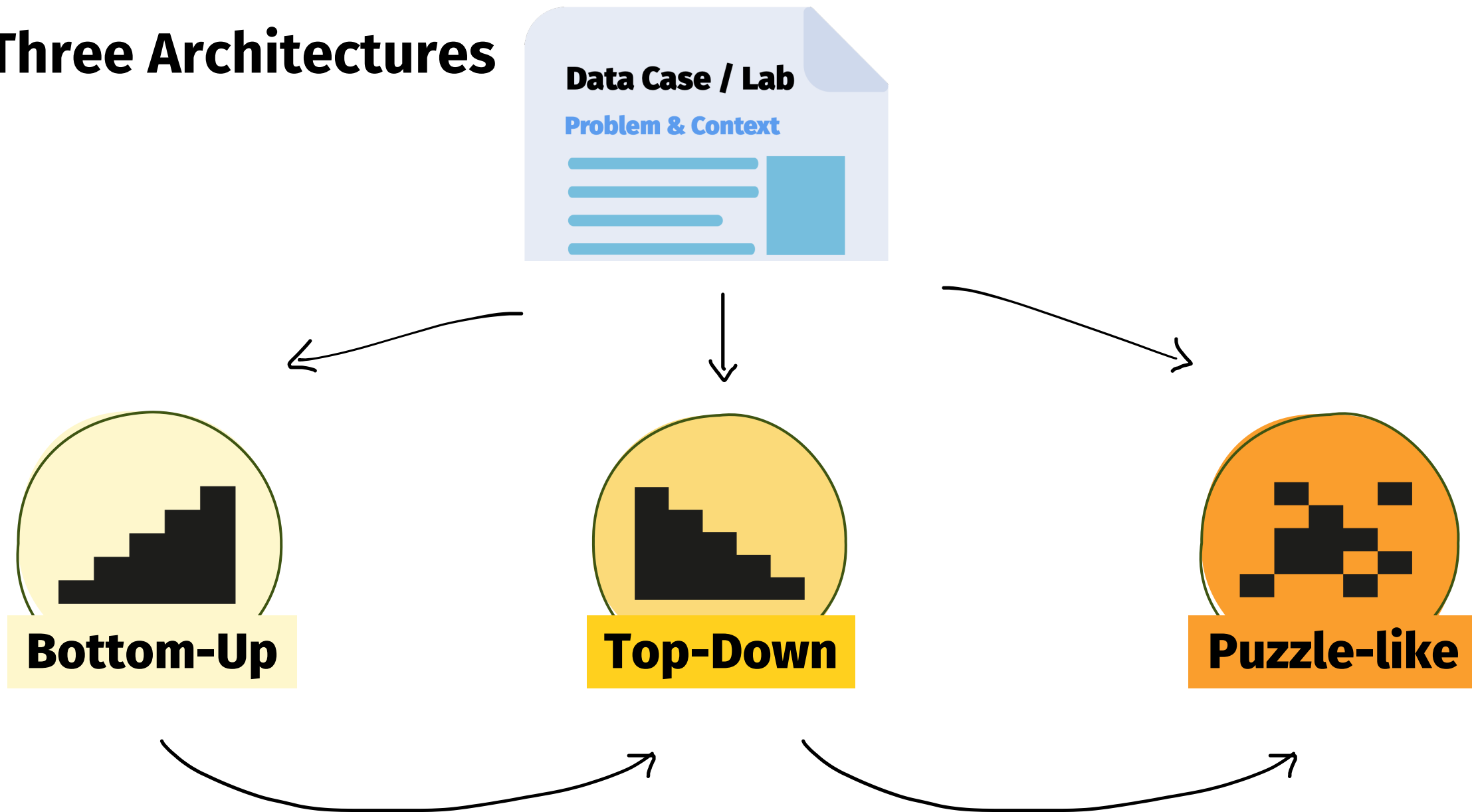
**School-specific data
case study in 3
architectures**

The school-specific data case study

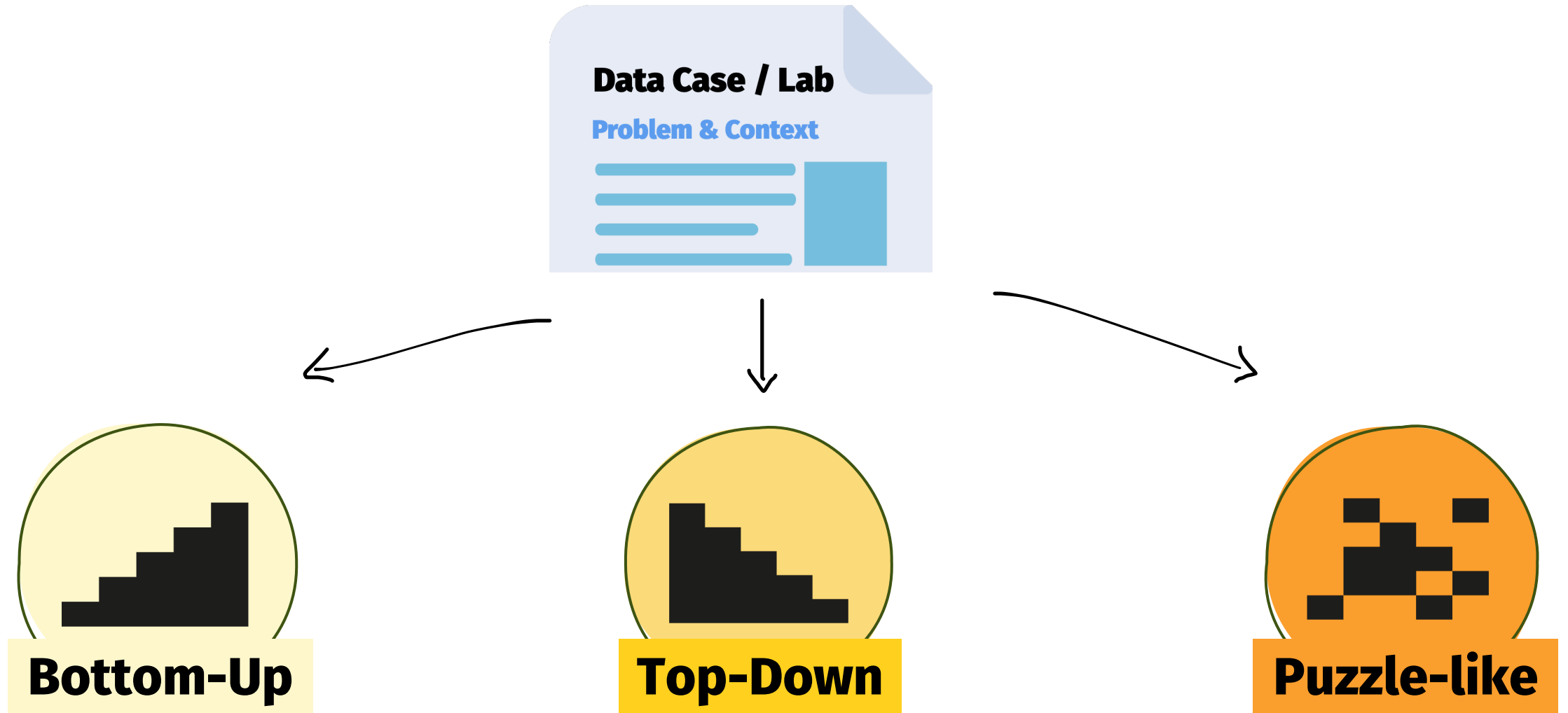
enables students using a real world data case



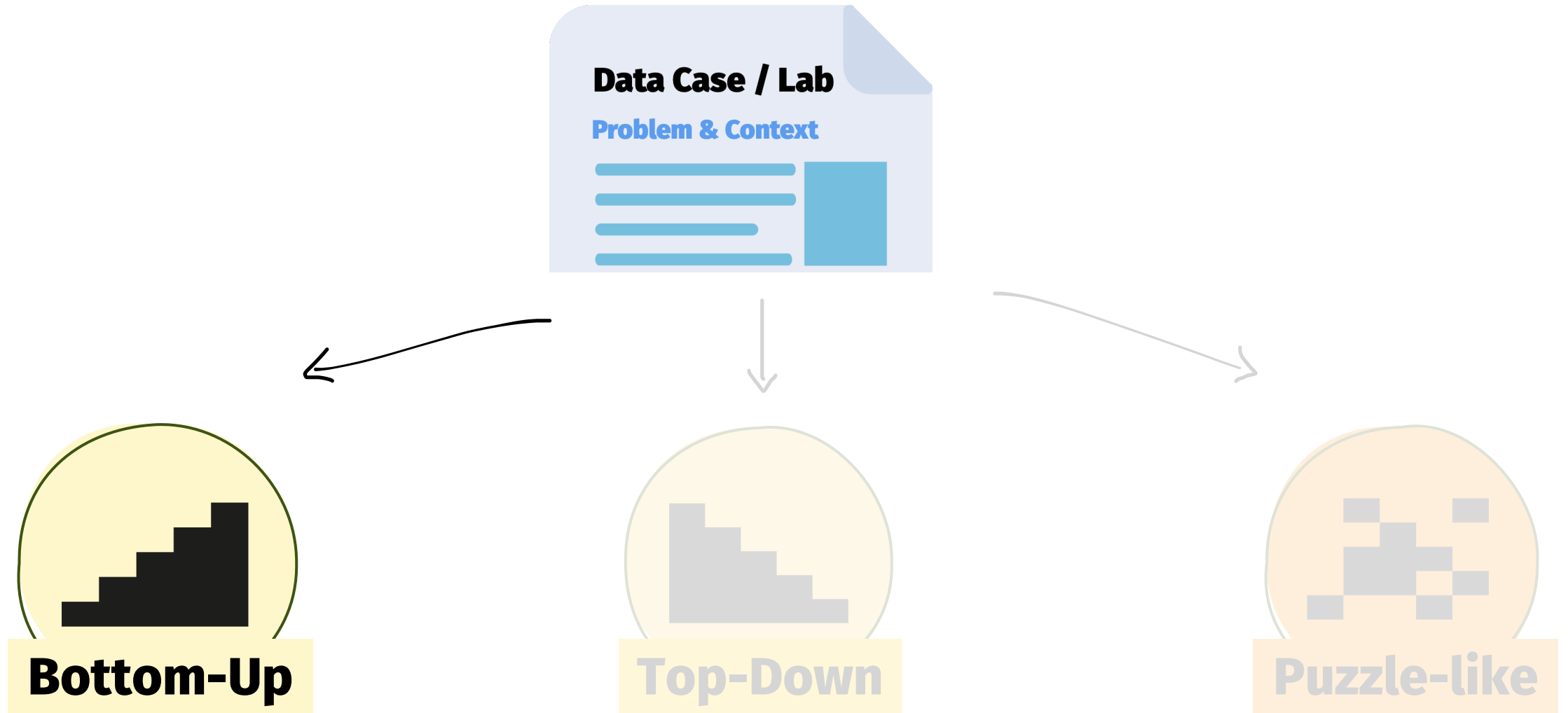
Three Architectures



Three Architectures



Three Architectures



The data case study

School-specific challenges in the transfer process

Lack of programming skills

Lack of domain knowledge

The time for preparing the data cases is limited.



Get an overview of the dataset.

Drag **Data Table** onto the workspace and connect the output of **Datasets** to the input of **Data Table**.

Look at the data on abalones using **Data Table** and fill in the gaps.

The data set includes measurements of _____ sea snails known as abalone and is organized into _____ organized into columns. The columns represent different _____ of the abalones. Translated into English, we have data on the following characteristics:

1. _____
2. _____
3. _____

Mit Daten zu 4177 Abalonen, erhoben an der Küste der australischen Bundesstaat Tasmanien



Data Table

	Rings	Sex	Length
1	15	M	0.455
2	7	M	0.350
3	9	F	0.530
4	10	M	0.440
5	7	I	0.330
6	8	I	0.425
7	20	F	0.530
8	16	F	0.545
9	9	M	0.475
10	19	F	0.550
11	14	F	0.525
12	10	M	0.430
13	11	M	0.490
14	10	F	0.535
15	10	F	0.470
16	12	M	0.500

Bottom-Up

Aufgabe 3. ⚡ Berichte deiner Nachbarin oder deinem Nachbarn von dem Umweltproblem, über das du gelesen hast. Frage deine Nachbarin oder deinen Nachbarn nach dem Umweltproblem in ihrem oder seinem Bericht. Notiere die Gründe für das Aussterben der Abalonen. Erledigt?

8:32 Heißes, angesäuertes Wasser, Überfischung ☐

Warum dürfen die Abalonen im Norden nicht gefangen werden? Einigt euch und notiert das Ergebnis.

Es gab keine Informationen über eine angemessene Größengrenze. ☐

Aufgabe 4. ⚡ Notiere, was über die Abalonen im Norden und im Süden bekannt ist. Erledigt?

Abalonen im Süden	Abalonen im Norden	
<u>Im Süden sind die Abalonen größer und schwerer.</u>	<u>Die Abalonen erreichen nicht die gesetzliche Mindestgröße (138mm).</u>	☐

Aufgabe 5. ⚡ Ihr habt einen Datensatz *Abalone* mit den Messungen von 4177 nördlichen Abalonen erhalten. Ihr schlägt vor, aus diesem Datensatz eine Mindestgröße für den Fang im Norden zu bestimmen, um den Norden für Taucher zu öffnen. Nennt vier Beteiligte, die an oder von dem Ergebnis eurer Arbeit betroffen sind.

8:37 Einheimische, Touristen, Tasmanische Abalonen-Taucher, Vereinigung, Abalone Verarbeiter/ Taucher ☐

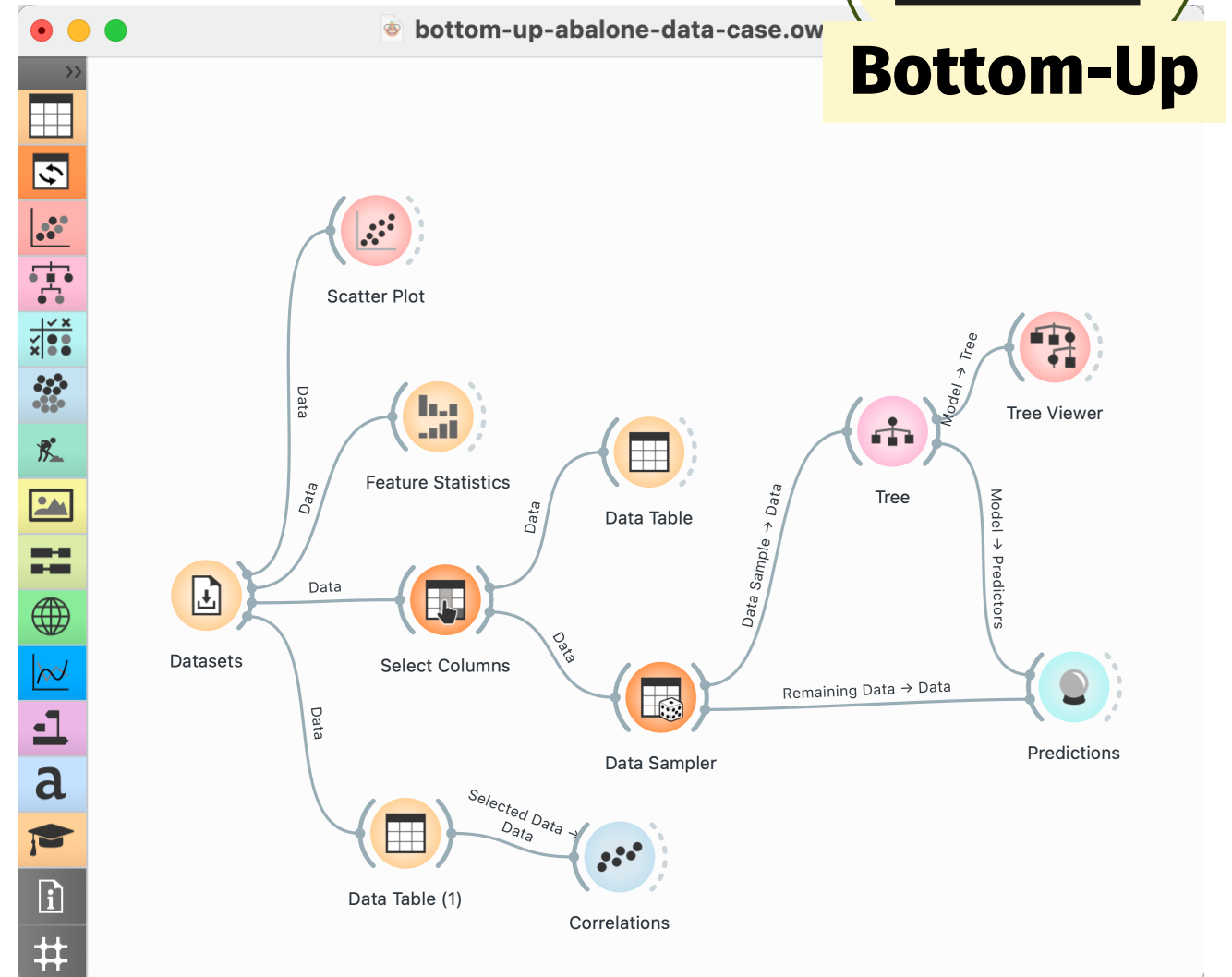
Aufgabe 6. ⚡ EINZELARBEIT: Nenne zwei beschreibende Fragen, die du untersuchen musst, um den Norden für Abalone-Taucher zu öffnen.

8:39 Wie hat sich die (Durchschnitts)größe/Anzahl der Abalonen in den letzten 10 Jahren entwickelt? ☐

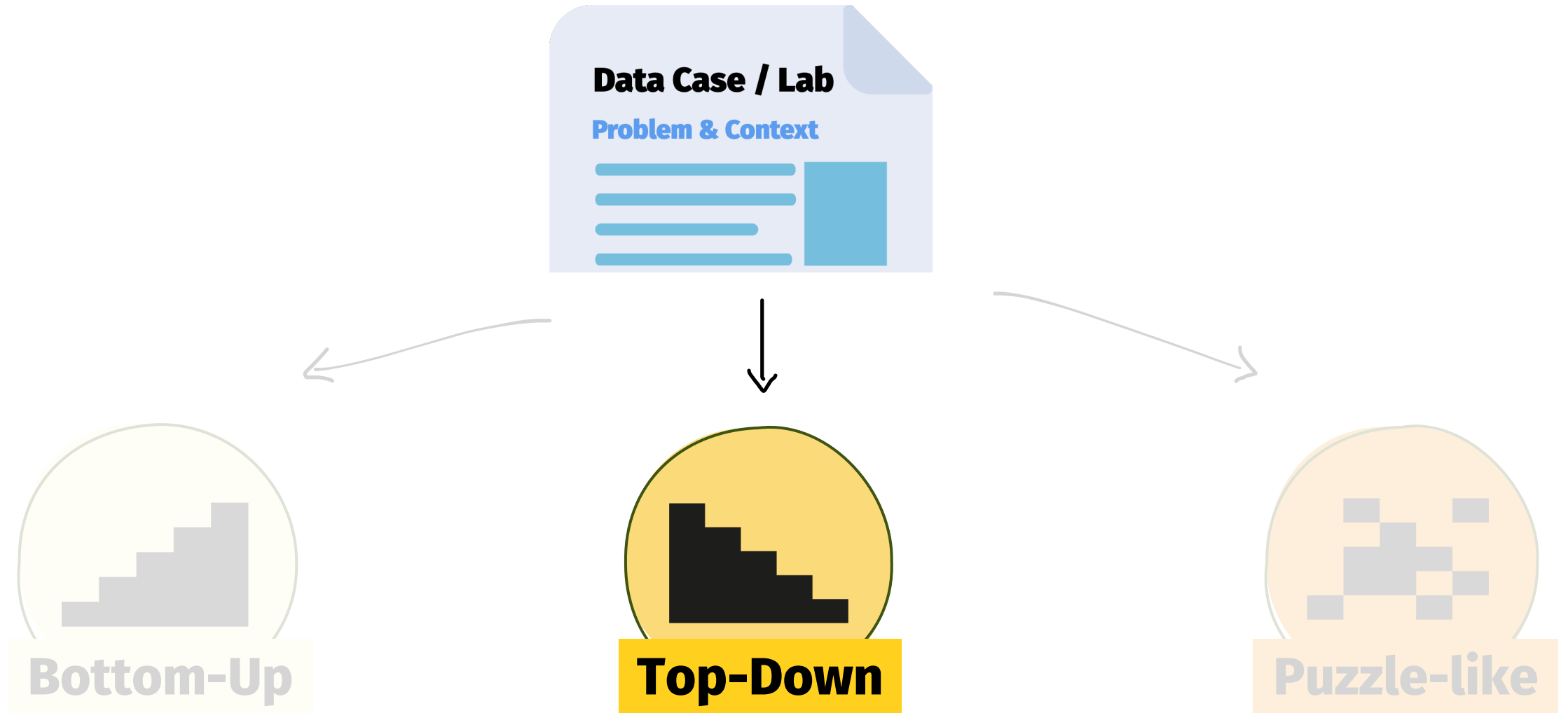
Aufgabe 7. ⚡ Vergleiche deine Antworten zu Aufgabe 6 mit denen deiner Nachbarin oder deines Nachbarn. Habt ihr die gleichen Schwerpunkte gesetzt? Erledigt?

8:41 ☐

Learning products are the **data flow** and the **written interpretation of findings** derived from the stages of the data-driven lifecycle.



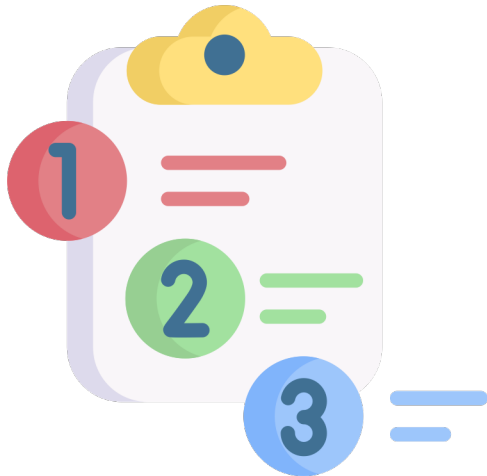
Three Architectures



Challenges



Overemphasis on the operating skills of Orange3



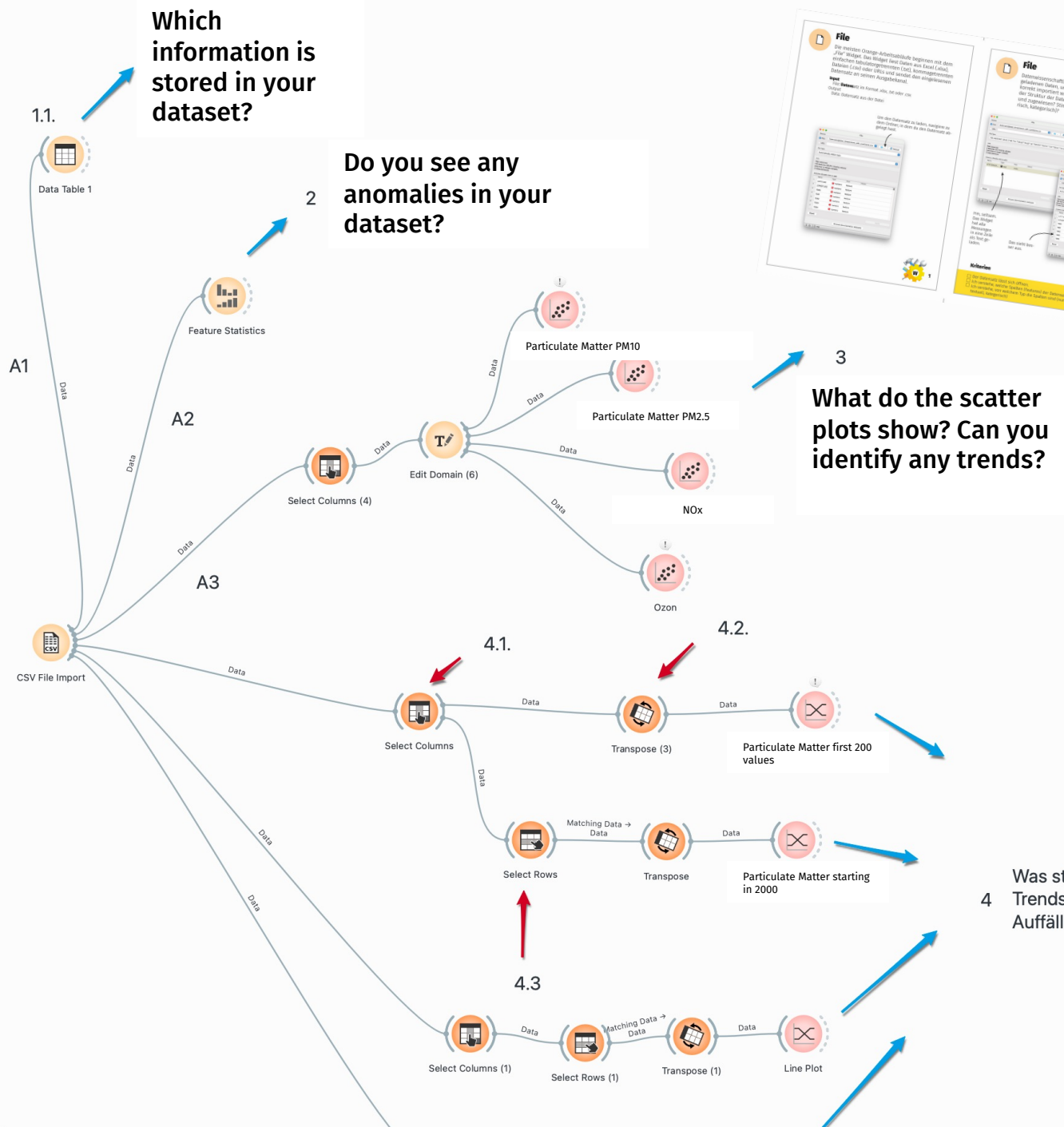
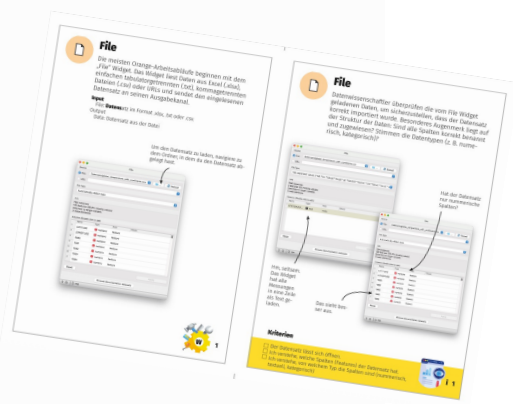
There is no reflection on the data-driven lifecycle.





Which information is stored in your dataset?

Do you see any anomalies in your dataset?



3. Examine the scatter plots. Summarize your impressions of the trend and any notable features.

A3: Give the branch a meaningful phase heading.

4 Was st Trends Auffäll

Interpretation of findings from the data-driven lifecycle (from the leaf nodes)

Description of the stages in the data-driven lifecycle



Top-Down

Investigate leaf widgets	Observations	Infokart
1. Lade den Datensatz in Orange3 und fasse deinen ersten Eindruck über die Daten zusammen.	* Daten fehlen * Koordinaten der Stationen vorhanden * nicht vertreten: Afrika, Antarktis, Grönland, Centralasien, wenige in hat * fehlende Daten am meisten im Jahr 1880, am wenigsten 1962 * Daten realistisch	w1, i1
2. Betrachte die Visualisierungen und fasse deinen zweiten Eindruck über die Daten zusammen (vor allem im Hinblick auf Auffälligkeiten).	- Es sind Daten zu verschiedenen Städten in den verschiedenen Jahren aufgelistet - in Form einer Tabelle (kein per se Daten) * vor allem Nordamerika und Europa * ausreißer in Südamerika * vereinzelt in Asien ein wenig in Australien	w2, i2
3. Überprüfe die Dokumentation. Fasse kurz zusammen, woher der Datensatz kommt.	* sind relevant * ursprünglich von NOAA und NASA	w3, i3
Vor dem Hintergrund der Analyse von 1-3 lies das Ziel noch einmal durch und begründe, ob es den SMART-Kriterien entspricht.	* nicht wirklich spezifisch * keine richtige Fragestellung * nicht wirklich messbar, aber attraktiv * realistisch * terminiert	w4, i4
4. Betrachte das Scatter Plot. Fasse deinen Eindruck über den Trend und die Auffälligkeiten zusammen.	* Jahrestemperaturen in Jahren in Form von Punkten * TRENDS * Auffälliger Anstieg der Temperaturen mit der Zeit	w5
5. Betrachte das Line Plot (und optional 6. Bar Plot). Fasse deinen Eindruck über den Trend und die Auffälligkeiten zusammen.	* ein Trend (Temperaturen in Jahren ab 1880 bis 2014) * Balkendiagramm: Kategorien werden verglichen (Abweichungen positiv oder negativ?)	
Vergleiche die Darstellungsarten in 4,5,6. Fasse ihre Stärken und Schwächen für die Analyse von Temperaturdaten zusammen.	* Scatter Plot: gut zum Ablesen einzelner Daten * Line Plot: gut um Verlauf zu erkennen * Bar Plot: gut um Abweichungen zu erkennen	
Fasse die Erkenntnisse aus der Trendanalyse in Bezug auf dein Ziel zusammen.	- Temperatur in Jahren von 1880 bis 2014 - aber mit schwankenden Temperaturen an, es die meisten sind vorhanden und sind	
Checkpoint 1:	Checkpoint 2:	
☒ Ich habe ein erstes Verständnis des Datensatzes. ☒ Ich weiß, ob der Datensatz Auffälligkeiten enthält. ☒ Der Datensatz ist relevant für meine Fragestellung.	☒ Ich habe die Frage vollständig teilweise gar nicht beantwortet ☒ Ich weiß, welche Analysen notwendig sind, um ausführlicher zu beantworten.	

Task Sheet 2: What is the meaning of the widgets in the branches, and what is the meaning of the branches for the data-driven process?

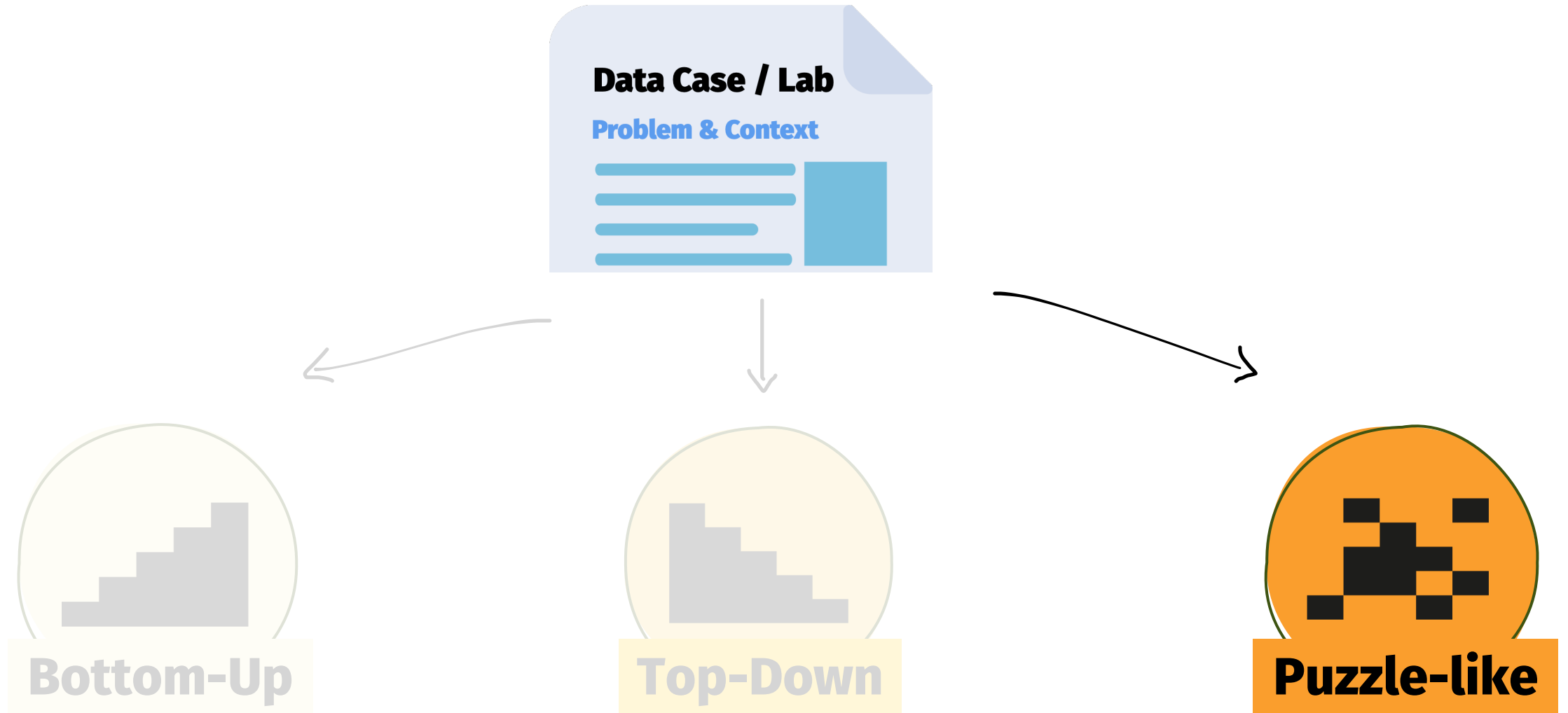
Gehe nun durch die Widgets in den Ästen (1-7) und beantworte die Fragen hinter den roten und violetten Pfeilen. Um den Zweck der einzelnen Widgets zu verstehen, lohnt es sich, einen Blick auf den Input und Output des Widgets sowie auf die Info zu werfen.

STATION	NAME	LATITUDE
40	CA00010000 BEATRICE, O.	45.1323
41	CA00010004 PARRY SOU...	46.1333
42	CA00010000 MONTREAL...	45.5
43	CA00010000 SOUTHAM...	44.5
44	CA00010000 HOHENPE...	47.8017

Investigate branch widgets

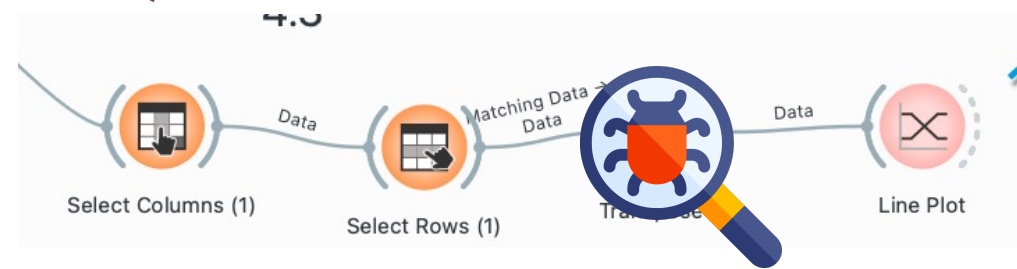
Investigate branch widgets	Observations
1.1 Warum werden die Daten in das Data Table Widget geladen?	für eine gute Übersicht und eventuelle Auswertung einzelner Spalten
A1 Gib eine aussagekräftige Phasenüberschrift für den Ast 1.	Das Übertragen der Daten in eine übersichtliche Tabelle
2.1 Wozu dient das Widget Select Rows an dieser Stelle?	Um die Daten in eine übersichtliche Tabelle zu übertragen
A2 Gib eine aussagekräftige Phasenüberschrift für den Ast 2.	Auswahl der Daten, die im Scatter-Widget dargestellt werden sollen
4.1 Wozu dient das Widget Select Columns an dieser Stelle?	Übertragung der Daten in eine geographische Karte
4.2 Wozu dient das Widget Transpose? Warum brauchen wir Transpose an dieser Stelle?	dient zur Auswahl der Spalten, die in den Scatter-Widget dargestellt werden sollen (hier: alle Jahre)
4.3 Wozu dient das Widget Edit Domain an dieser Stelle?	dient zur Festlegung/Bearbeitung des Orts, den Daten eingezeichnet werden sollen
4.4 Wozu dient das Widget Preprocess an dieser Stelle?	Um festzulegen, welche Daten genommen werden
A3 Gib eine aussagekräftige Phasenüberschrift für die Aste 4-6.	beantwortet die Frage, was mit den fehlenden Daten gemacht werden/dargestellt werden sollen
	Das Auswählen der Daten und das letztendliche Visualisieren dieser

Three Architectures

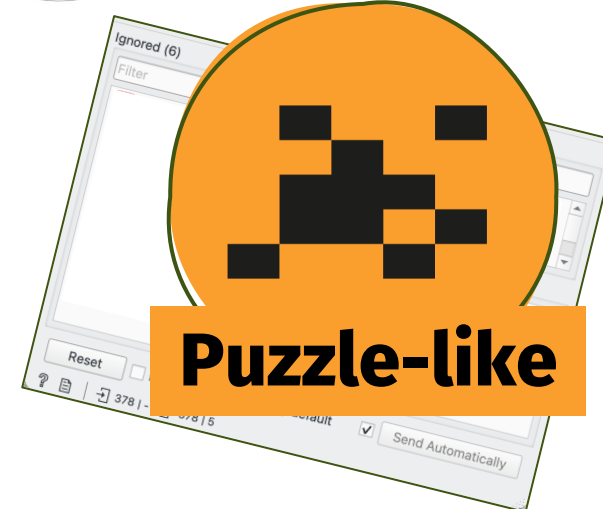
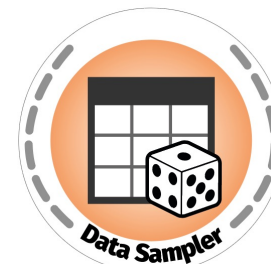
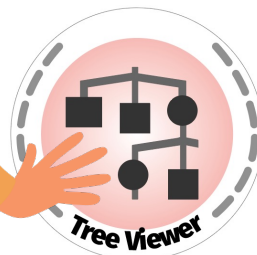
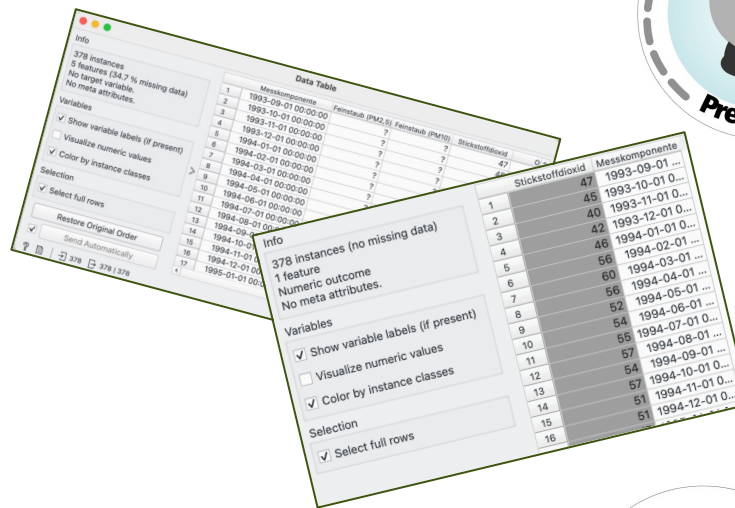
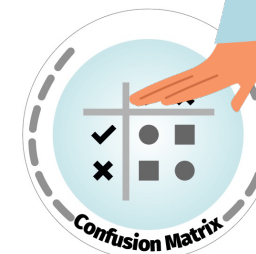
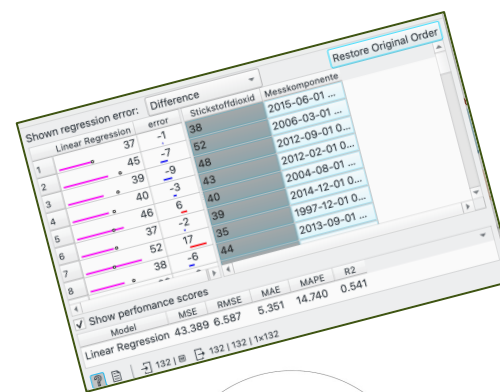
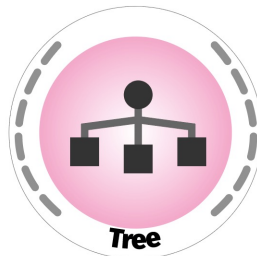
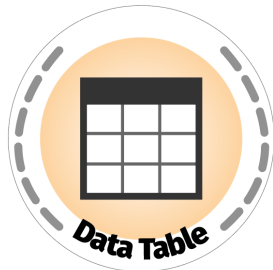


Challenges

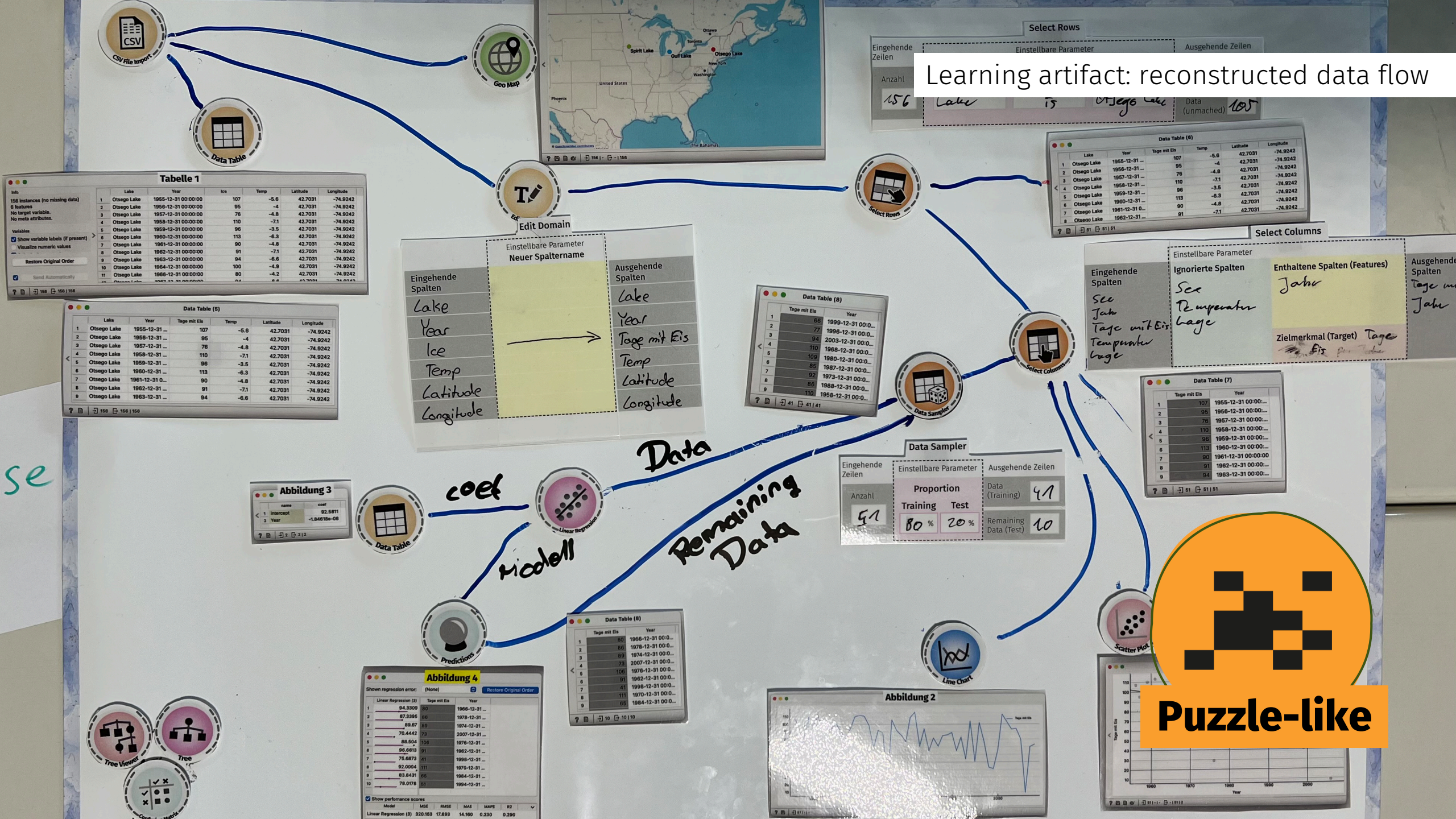
Data case studies **offer few opportunities for exchange**, “Where I really need to say something to someone else, where I really (...) get into a real conversation.” (S1T1W3)



Students lack basic debugging skills.

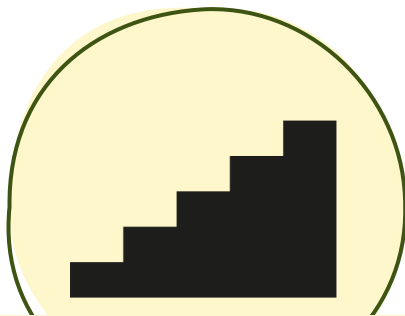
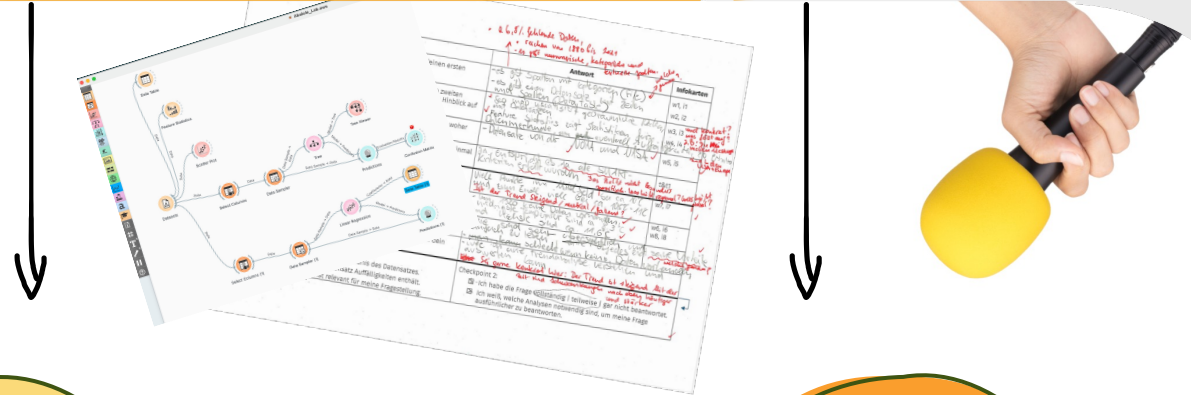


Learning artifact: reconstructed data flow





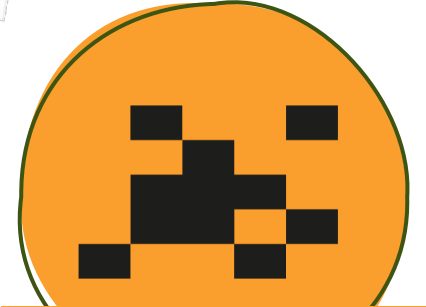
Summary



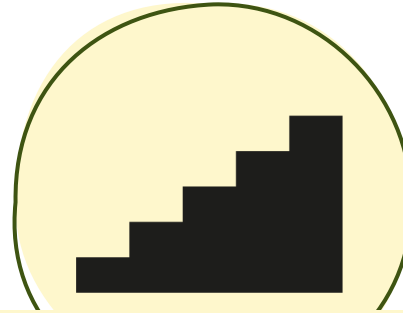
Bottom-Up



Top-Down



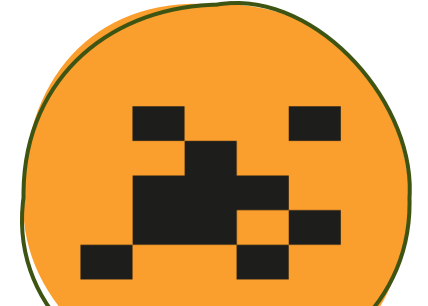
Puzzle-like



Bottom-Up



Top-Down



Puzzle-like

Challenges

Lack of programming experience	Yes	Yes	Yes
Reflection on the process	No	Yes	Yes
Lack of debugging skills	No	No	Yes
...	...	...	...
Data concepts and practices	Introduction to a data practice or data concept	Phases in the data flow, reflection on this	Data flow, debugging flows

<https://computingeducation.de>



RESEARCH-ARTICLE



The Data Case Study - A Teaching and Learning Method for Computer Science Education in Schools

Authors:  [Viktoriya Olari](#),  [Ralf Romeike](#) | [Authors Info & Claims](#)

[CompEd 2025: Proceedings of the ACM Global on Computing Education Conference 2025 Vol 1](#) • Pages 225 - 232
<https://doi.org/10.1145/3736181.3747164>

Published: 21 October 2025 [Publication History](#)



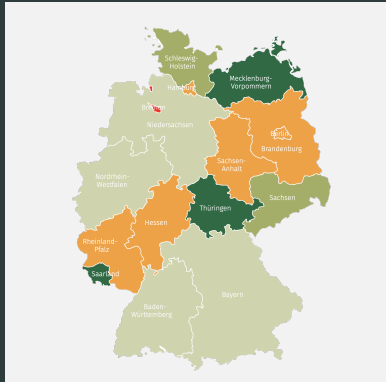
Can be used as a foundation for a common language when teaching about data-driven systems.

Can be used as a guide when specifying competencies.

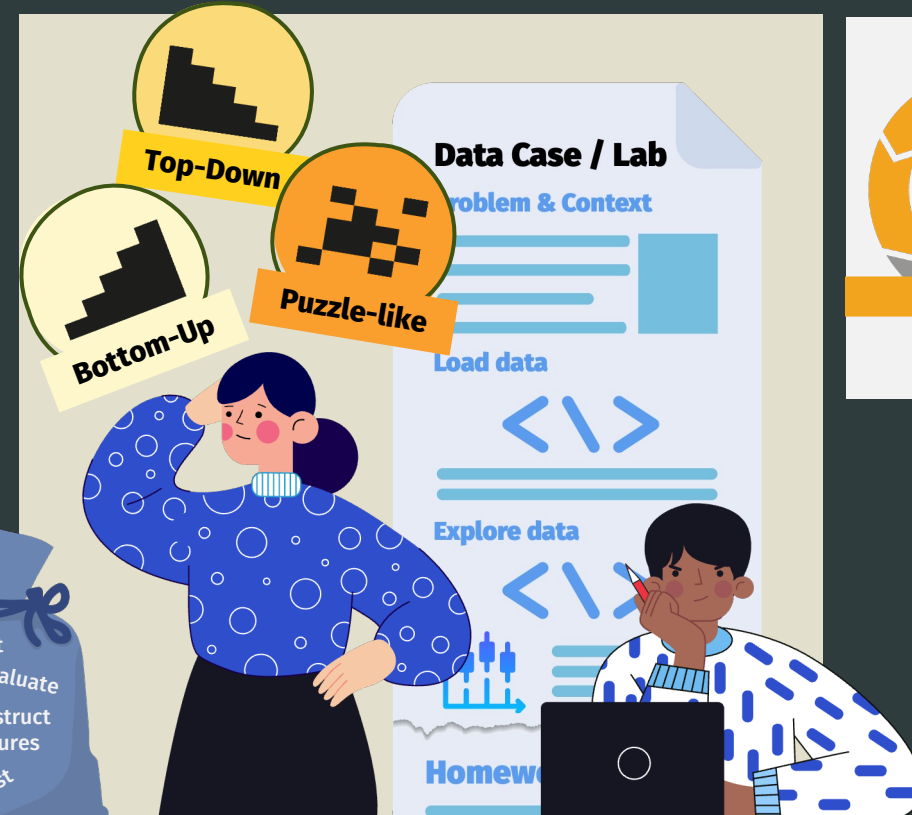
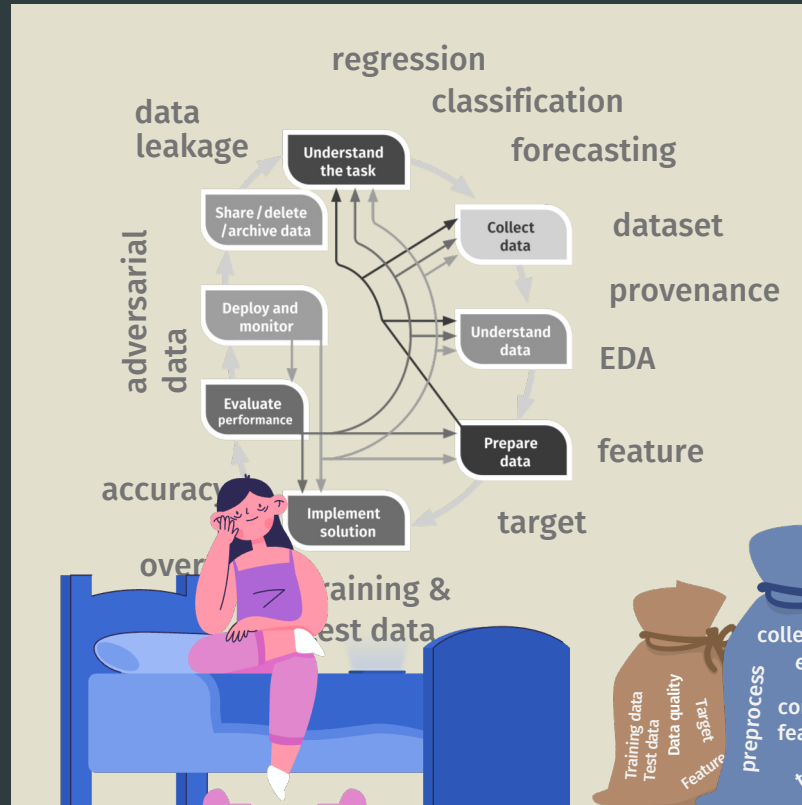
Adapting a method takes time and is challenging. Along the way, we have discovered new ways to teach about data.

Data concepts & practices

Context & Motivation

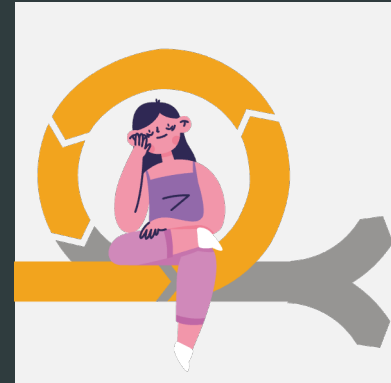


↓
AI content is entering formal education.



The data case study

Summary



Sources

Core curriculum for secondary level I – grades 5 and 6 in North Rhine-Westphalia (2021), https://lehrplannavigator.nrw.de/system/files/media/document/file/si_kl5u6_if_klp_2021_07_01.pdf.

Peter J. Denning. 2003. Great principles of computing. Commun. ACM 46, 11 (November 2003), 15–20. <https://doi.org/10.1145/948383.948400>

Andreas Schwill. 1994. Fundamental Ideas of Computer Science. (1994). Retrieved from <http://ddi.cs.uni-potsdam.de/Forschung/Schriften/EATCS.pdf>

David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What Should Every Child Know about AI? In Proceedings of the AAAI Conference on Artificial Intelligence, July 17, 2019. 9795–9799. <https://doi.org/10.1609/aaai.v33i01.33019795>

Matti Tedre, Peter Denning, and Tapani Toivonen. 2021. CT 2.0. In Proceedings of the 21st Koli Calling International Conference on Computing Education Research (Koli Calling '21), November 2021. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3488042.3488053>

Tilman Michaeli, Ralf Romeike, and Stefan Seegerer. 2022. What Students Can Learn About Artificial Intelligence – Recommendations for K-12 Computing Education. In Towards a Collaborative Society Through Creative Learning, 2022. Springer Nature Switzerland, Cham, 196–208. https://doi.org/10.1007/978-3-031-43393-1_19

Jane Waite, Ethel Tshukudu, Veronica Cucuiat, Robert Whyte, and Sue Sentance. 2023. Towards a Framework for Learning Content Analysis in K-12 AI/ML Education. In 2023 IEEE Frontiers in Education Conference (FIE), October 18, 2023. IEEE, College Station, TX, USA, 1–5. <https://doi.org/10.1109/FIE58773.2023.10343368>

Viktoriia Olari and Ralf Romeike. 2021. Addressing AI and Data Literacy in Teacher Education: A Review of Existing Educational Frameworks. 2021. Association for Computing Machinery, Virtual Event, Germany, Article 17. <https://doi.org/10.1145/3481312.3481351>

Daochen Zha, Zaid Pervaiz Bhat, Kwei-Herng Lai, Fan Yang, and Xia Hu. 2023. Data-centric AI: Perspectives and Challenges. In Proceedings of the 2023 SIAM International Conference on Data Mining (SDM), Shashi Shekhar, Zhi-Hua Zhou, Yao-Yi Chiang and Gregor Stiglic (eds.). Society for Industrial and Applied Mathematics, Philadelphia, PA, 945–948. <https://doi.org/10.1137/1.9781611977653>

Johannes Jakubik, Michael Vössing, Niklas Kühl, Jannis Walk, and Gerhard Satzger. 2024. Data-Centric Artificial Intelligence. Bus Inf Syst Eng (March 2024). <https://doi.org/10.1007/s12599-024-00857-8>

Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. 2023. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Information Fusion 99, (November 2023), 101805. <https://doi.org/10.1016/j.inffus.2023.101805>

Kostas Kampourakis. 2018. On the Meaning of Concepts in Science Education. Sci & Educ 27, 7–8 (October 2018), 591–592. <https://doi.org/10.1007/s11191-018-0004-x>

Viktoriia Olari and Ralf Romeike. 2024. Data-related concepts for artificial intelligence education in K-12. Computers and Education Open 7, (December 2024), 100196. <https://doi.org/10.1016/j.caeo.2024.100196>

Sandro Costa and Fabiele Broietti. 2023. What are scientific practices and why are they relevant in science education? Revista Brasileira de Ensino de Ciência e Tecnologia 16, (January 2023), 1–22. <https://doi.org/10.3895/rbect.v16n1.14897>

Sources (continued)

Viktoriya Olari and Ralf Romeike. 2024. Data-related practices for creating Artificial Intelligence systems in K-12. Proceedings of the 19th WiPSCE Conference on Primary and Secondary Computing Education Research, Article 5. Retrieved from <https://doi.org/10.1145/3677619.3678115>

OECD. 2025. Empowering learners for the age of AI: An AI literacy framework for primary and secondary education (Review draft). OECD, Paris. Retrieved from <https://ailiteracyframework.org>

Orit Hazzan and Koby Mike. 2024. Guide to teaching data science: an interdisciplinary approach. Springer, Cham.

Yim Register and Amy J. Ko. 2020. Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics. In Proceedings of the 2020 ACM Conference on International Computing Education Research, August 10, 2020. ACM, Virtual Event New Zealand, 67–78. <https://doi.org/10.1145/3372782.3406252>

Chantel Ridsdale, James Rothwell, Mike Smit, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, Brad Wuetherick, and Hossam Ali-Hassan. 2015. Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report. (2015). <https://doi.org/10.13140/RG.2.1.1922.5044>

Deborah Ann Nolan and Terry Speed. 2001. Stat labs: mathematical statistics through applications (Corr. 2. print ed.). Springer, New York Berlin Heidelberg.

Valentina Chkoniya. 2021. Success Factors for Using Case Method in Teaching Applied Data Science Education. European Journal of Education 4, 1 (April 2021), 77–86. <https://doi.org/10.26417/236hbm84v>

Carrie Wright, Qier Meng, Michael R. Breshock, Lyla Atta, Margaret A. Taub, Leah R. Jager, John Muschelli, and Stephanie C. Hicks. 2024. Open Case Studies: Statistics and Data Science Education through Real-World Applications. Journal of Statistics and Data Science Education (2024), 1–30. <https://doi.org/10.1080/26939169.2024.2394541>

Erin Rae Hoffer. 2020. Case-based Teaching: Using Stories for Engagement and Inclusion. 2, 2 (2020), 75–80.

T. Grandon Gill. 2011. Informing with the case method: a guide to case method research, writing, & facilitation. Informing Science Press, Santa Rosa.

Clyde Freeman Herreid. 2011. Case study teaching. New Drctns for Teach & Learn 2011, 128 (December 2011), 31–40. <https://doi.org/10.1002/tl.466>

Jana Lasser, Debsankha Manik, Alexander Silbersdorff, Benjamin Säfken, and Thomas Kneib. 2021. Introductory data science across disciplines, using Python, case studies, and industry consulting projects. Teaching Statistics 43, S1 (July 2021). <https://doi.org/10.1111/test.12243>

Claire Toogood. 2023. Supporting students to engage with case studies: a model of engagement principles. Educational Review (November 2023), 1–15. <https://doi.org/10.1080/00131911.2023.2281227>

Susanne Prediger, Michael Link, Renate Hinz, Stephan Hussmann, Bernd Ralle, and Jörg Thiele. 2012. Lehr-Lernprozesse initiieren und erforschen. Fachdidaktische Entwicklungsforschung im Dortmunder Modell. Der mathematische und naturwissenschaftliche Unterricht 65, 8 (2012), 452–457.

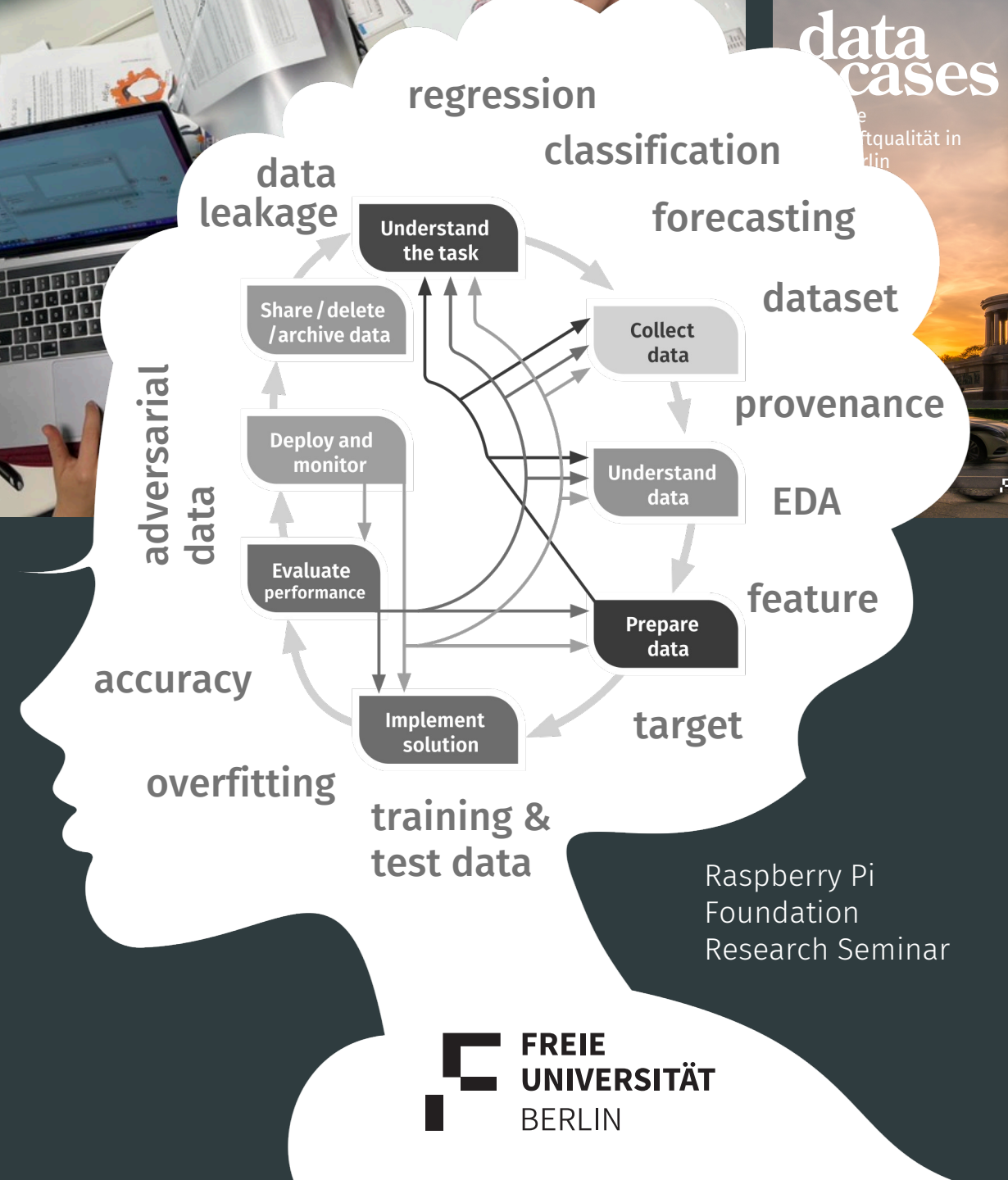


Data-related concepts and practices for AI education in K-12

Viktoriya Olari | 14.10.2025



Computing Education
Computer science explained!



Raspberry Pi
Foundation
Research Seminar

**FREIE
UNIVERSITÄT
BERLIN**