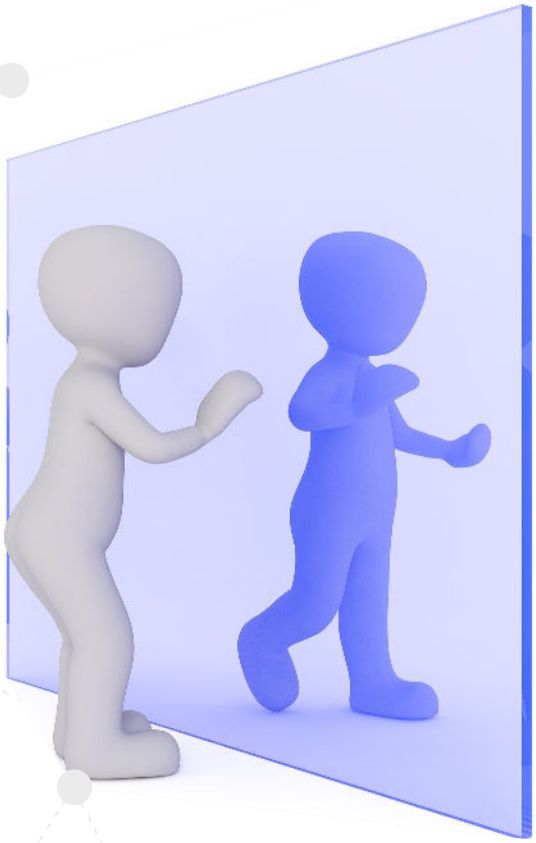
PADERBORN UNIVERSITY

Deutsche Telekom Stiftung



Paderborn Colloquium on Artificial Intelligence and Data Science Education at School Level, 18th May 2022

Data Awareness

Be Aware of the Data!

Carsten Schulte und **Lukas Höper**
Computing Education at Paderborn University in Germany

Project Data Science and Big Data at School (ProDaBi)

Initiated and funded by the Deutsche Telekom Stiftung

ProDaBi I : 2018 - 2020
ProDaBi II: 2020 - 2023

Cooperation at the Paderborn University:

Rolf Biehler	Didactics of Mathematics
Carsten Schulte	Didactics of Computer Science


Associates:

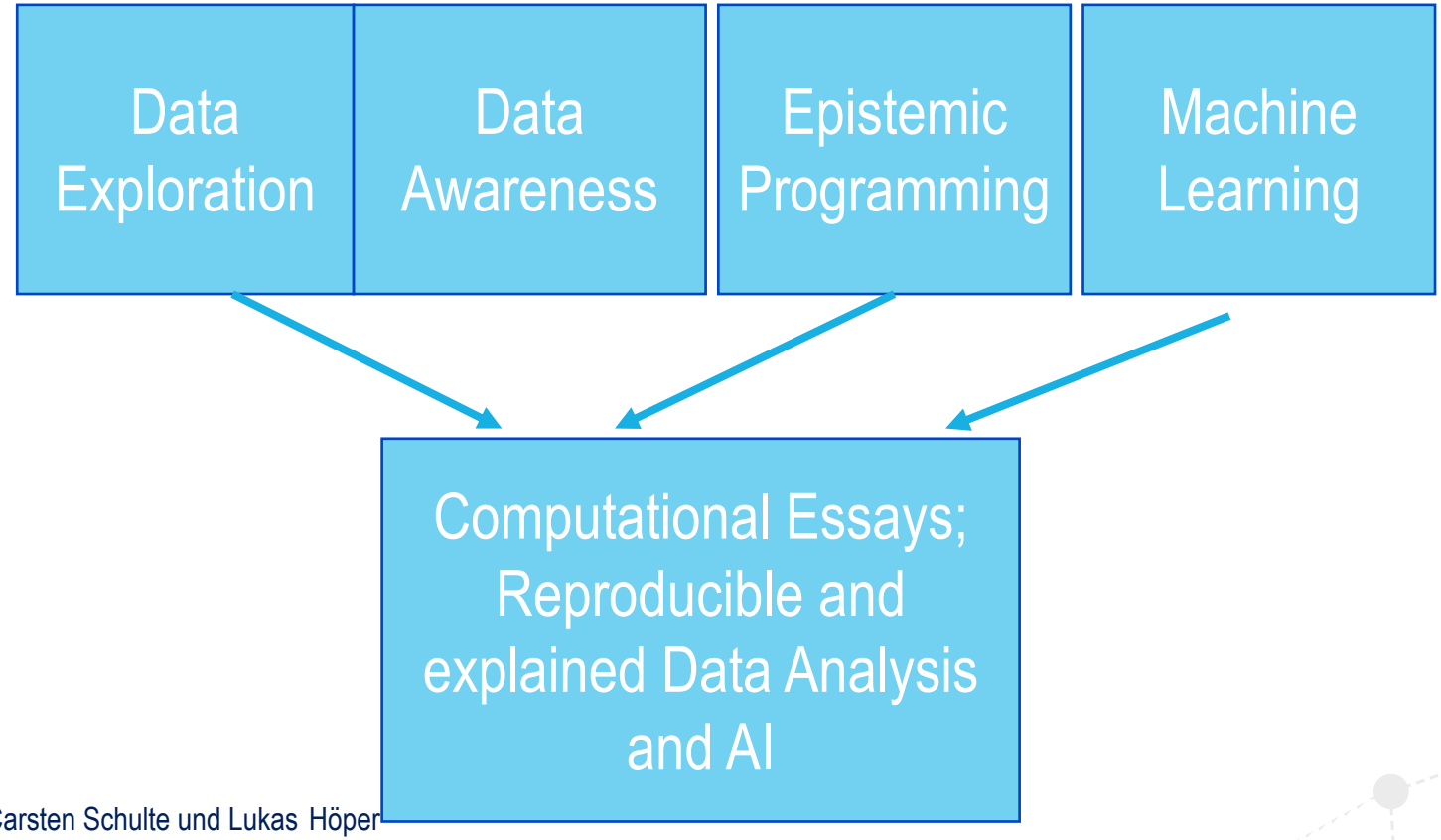
Yannik Fleischer	Didactics of Mathematics
Daniel Frischemeier	Didactics of Mathematics
Lukas Höper	Didactics of Computer Science
Sven Hüsing	Didactics of Computer Science
Susanne Podworny	Didactics of Mathematics





www.prodabi.de/en

Main topics and concepts of ProDaBi



Context

Project Course 12

- Data Project Topic
- JIM-Data (CODAP)
- Introduction to Python
- Data Exploration (JN)
- Decision Trees
- Neural Nets
- Project

Modules 8-10

- JIM-Data (CODAP)
- Data Exploration (JN, Arduino)
- Decision Trees (CODAP, JN)
- Data Awareness

Grade 7

An Initiative of the Federal Ministry of Education and Research

Science Year 2019

ARTIFICIAL INTELLIGENCE

Modules 5 / 6

- Decision Trees & Data Cards
- Data Awareness

ProDaBi material in different grades

Data Science in grade 12
Yearlong „project course“

Data Science in grade 8 to 10
5 teaching modules + PD courses for teachers

Data Science in grade 5 and 6
2 teaching modules

Teaching unit:
Cellular network and collection of location data

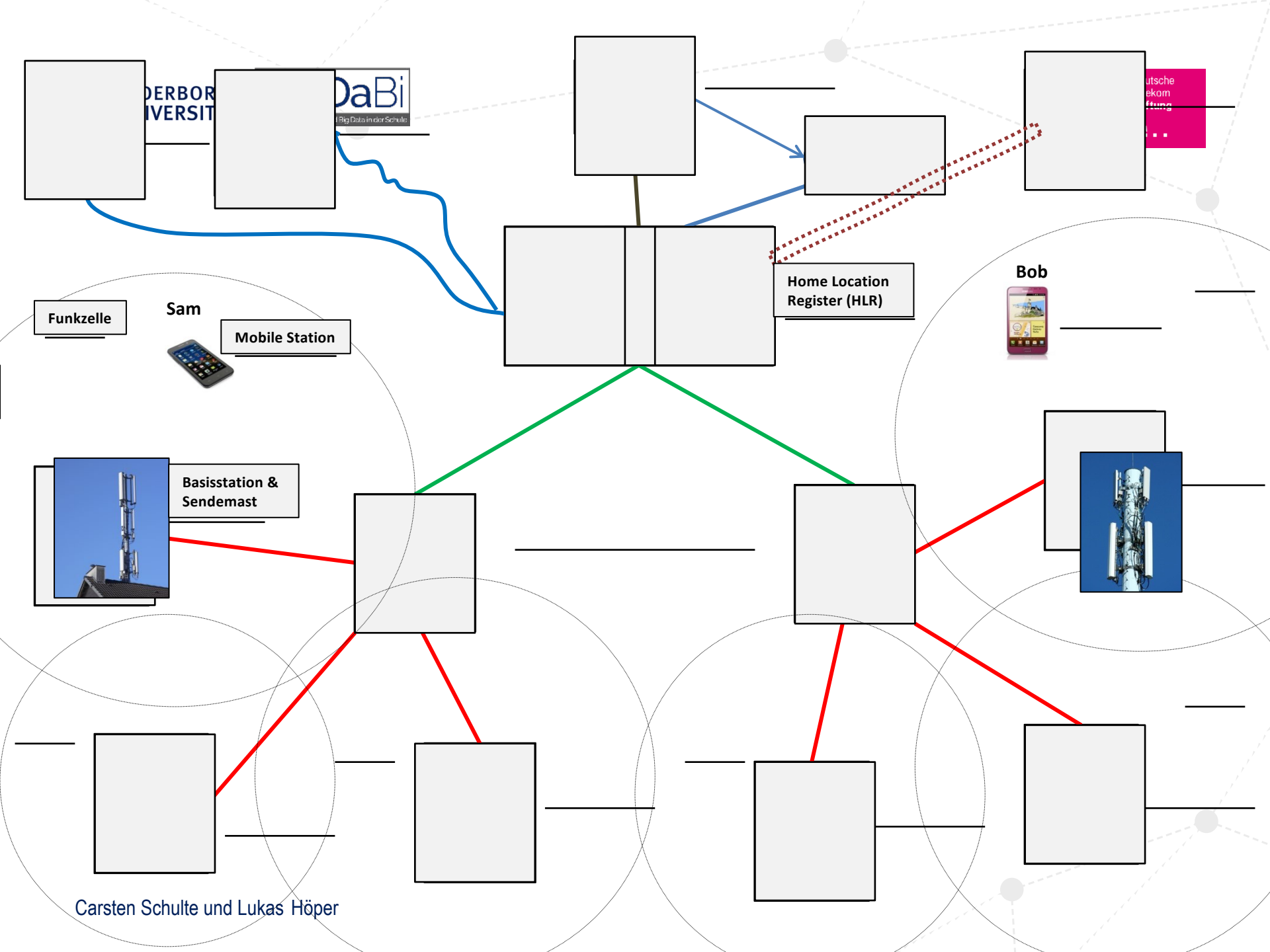
(old version)

Privacy & the Mobile Phone Network

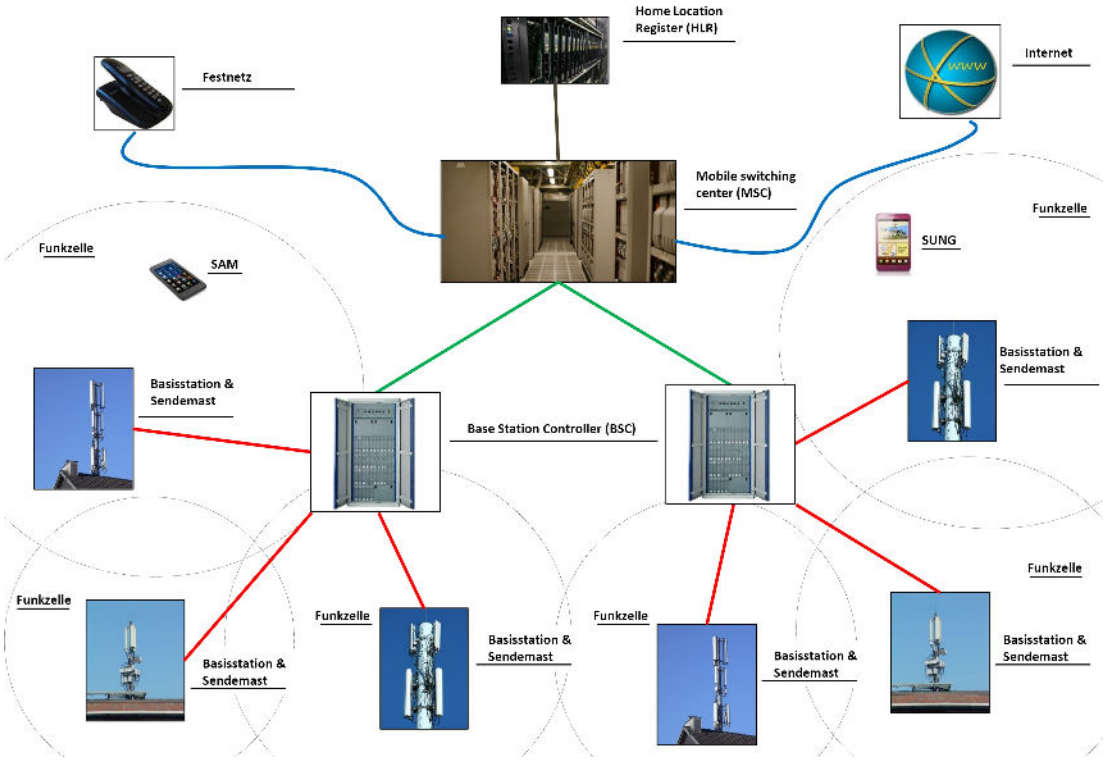
~ understanding digital artifacts
~ learning computer science (here e.g. informatics systems, protocols)
~ include societal impact (here e.g. privacy, data retention)

-> Does a data driven / data science education perspective change something?

The ‘old’ teaching unit

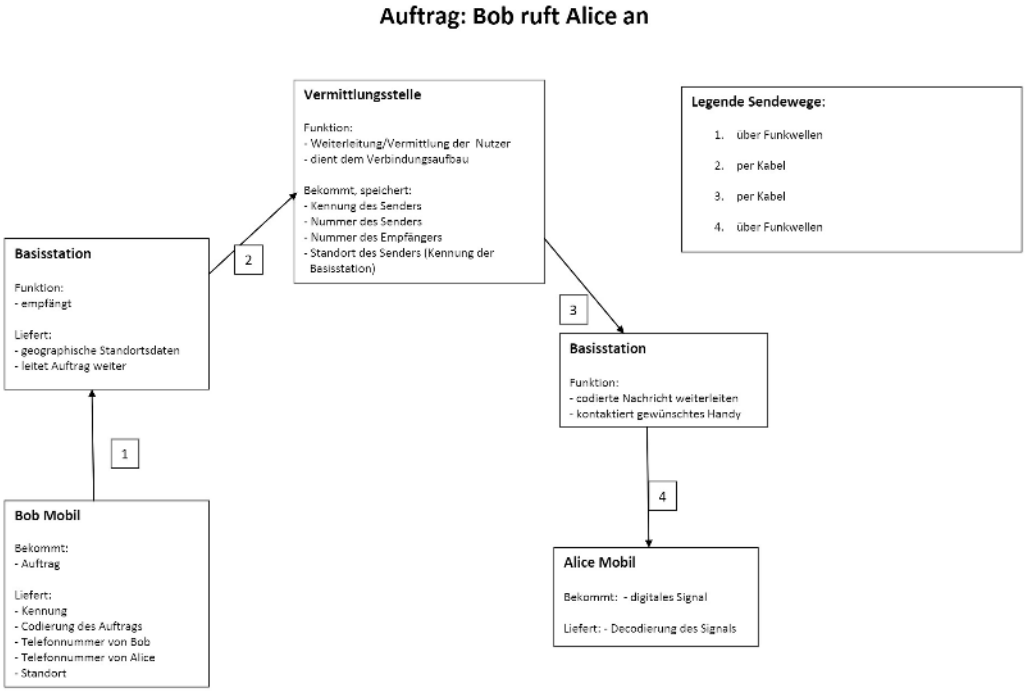

Home Location Register (HLR)
Bob
Funkzelle
Sam
Mobile Station
Basisstation & Sendemast
Carsten Schulte und Lukas Höper

The ‘old’ teaching unit



https://git.imp.fu-berlin.de/ddi/unterrichtsmodellmobilfunknetz/-/blob/master/PhaseEins/Gruppenpuzzle/1-Netzstrukturpuzzle-Musterl%C3%B6sung.pdf

The ‘old’ teaching unit



https://github.com/BastiSeitz/Ortsbezogene-Daten-im-Mobilfunknetz/blob/master/Lernabschnitt%201%20Struktur%20und%20Funktion%20des%20Mobilfunknetzes/Material%20und%20Hinweise%20zur%20Unterrichtsdurchf%C3%BChrung/Tafelbild.pdf

Data Protection

Betrayed by our own data

Mobile phones are tracking devices that reveal much about our lives. One look at our interactive map of data provided by the Green party politician Malte Spitz shows why.

Von **Kai Biermann**

10. März 2011, 17:09 Uhr / 2 Kommentare /

https://www.zeit.de/digital/datenschutz/2011-03/data-protection-malte-spitz

The ‘old’ teaching unit

MI.Lab
Freie Universität Berlin

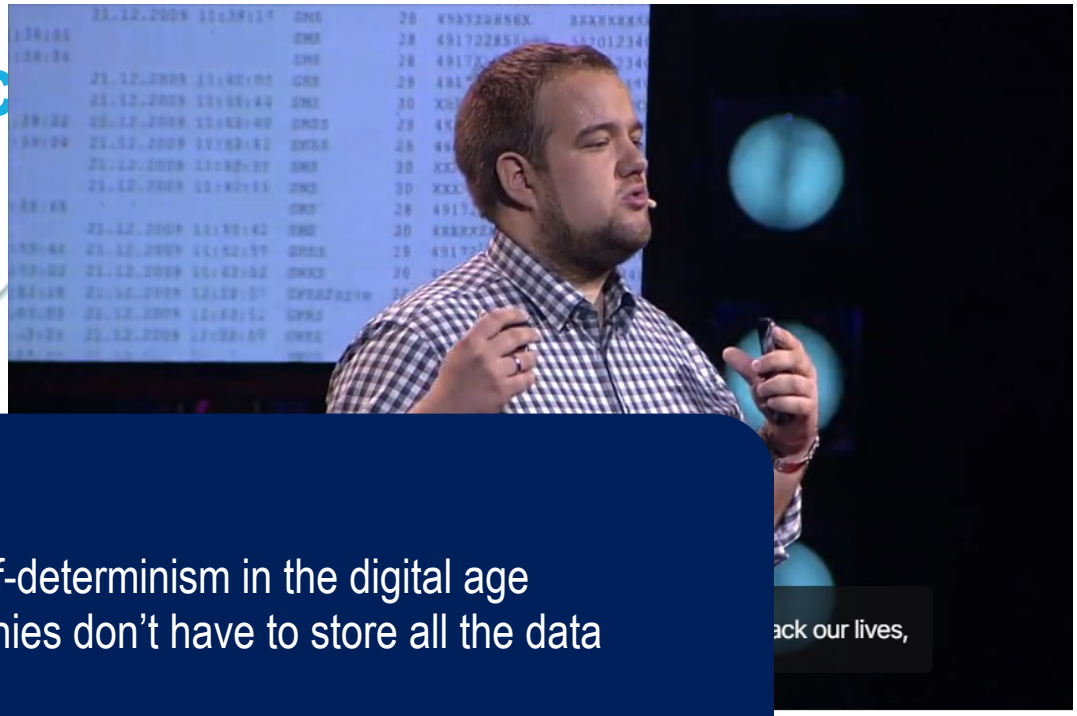
Project Profiler-Teams

Wir haben einen Mobilfunkdatensatz einer Person erhalten. Dieser erstreckt sich über den Zeitraum eines halben Jahres. Eure Aufgabe ist es in 3er Gruppen ein Profil dieser Person zu erstellen! Eurer Kreativität sind dabei keine Grenzen gesetzt!

Aufgabe: Das Profil der Person soll von Ihnen als Plakat angefertigt werden. Am Ende des Profiler-Projects müssen Sie ihr erstelltes Profil der Person präsentieren.

Folgendes bevor Sie mit Hilfe von Processing das „profilen“ beginnen:

The ‘old’ teac



Self-determinism in the digital age
Companies don’t have to store all the data

Durchl

Im Folge

Speicher

Entscheide

"Was sich aus unseren Einkäufen
2014). http://crackedlabs.org/stu

ack our lives,

Like (52K)

Sh

Your phone company is watching

What kind of data is your cell phone company collecting? Malte Spitz wasn't too worried when he asked his operator in Germany to share information stored about him. Multiple unanswered requests and a lawsuit later, Spitz received 35,830 lines of code -- a detailed, nearly minute-by-minute account of half a year of his life.

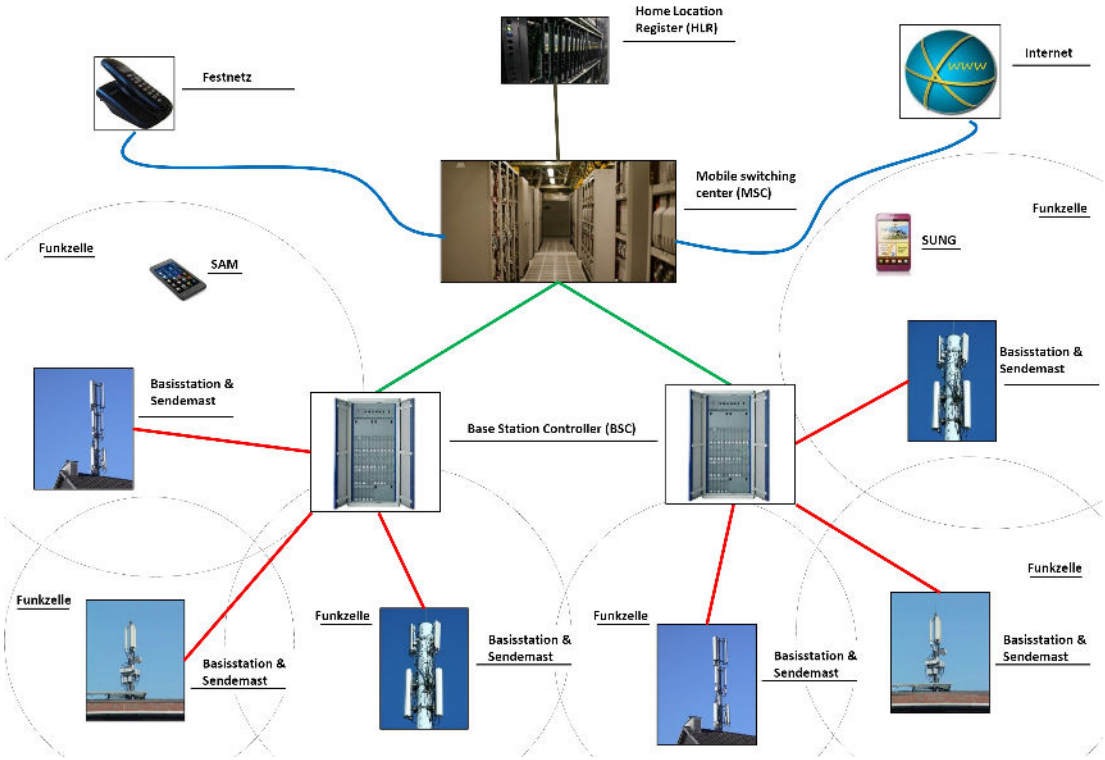
Privacy & the Mobile Phone Network

~ understanding digital artifacts
~ learning computer science (here e.g. informatics systems, protocols)
~ include societal impact (here e.g. privacy, data retention)

? Focus on threats in the old teaching unit & implicitly maybe teaching students how to behave instead of giving them tool to develop an informed opinion?

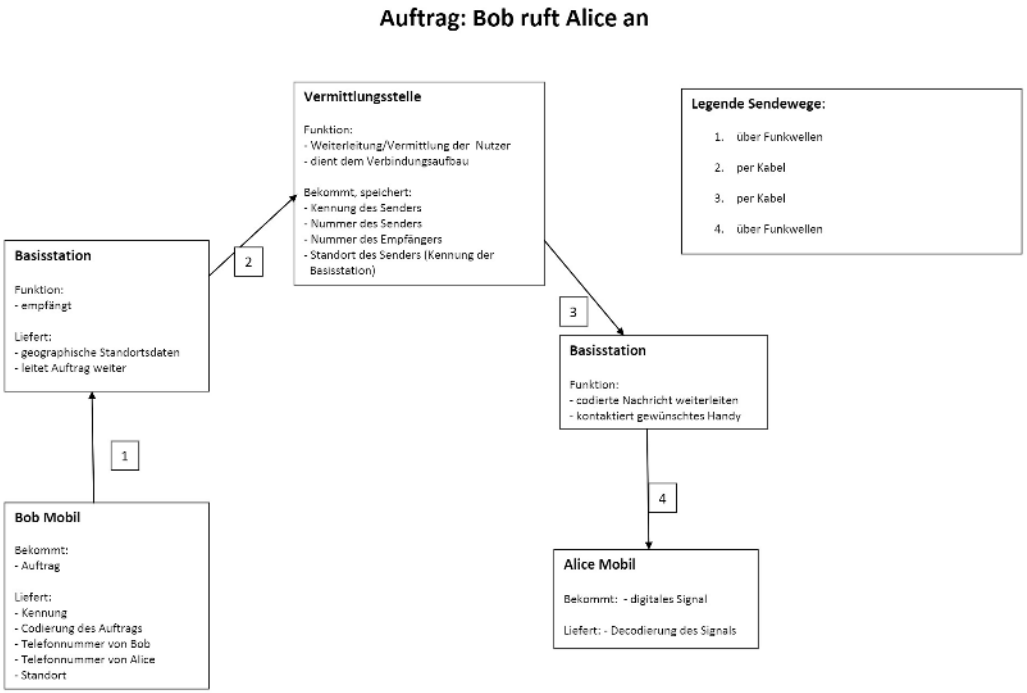
-> Does a data driven / data science education perspective change something?

The ‘old’ teaching unit



https://git.imp.fu-berlin.de/ddi/unterrichtsmodellmobilfunknetz/-/blob/master/PhaseEins/Gruppenpuzzle/1-Netzstrukturpuzzle-Musterl%C3%B6sung.pdf

The ‘old’ teaching unit



https://github.com/BastiSeitz/Ortsbezogene-Daten-im-Mobilfunknetz/blob/master/Lernabschnitt%201%20Struktur%20und%20Funktion%20des%20Mobilfunknetzes/Material%20und%20Hinweise%20zur%20Unterrichtsdurchf%C3%BChrung/Tafelbild.pdf

Data Protection

Betrayed by our own data

Mobile phones are tracking devices that reveal much about our lives. One look at our interactive map of data provided by the Green party politician Malte Spitz shows why.

Von **Kai Biermann**

10. März 2011, 17:09 Uhr / 2 Kommentare /

https://www.zeit.de/digital/datenschutz/2011-03/data-protection-malte-spitz

The ‘old’ teaching unit

MI.Lab
Freie Universität Berlin

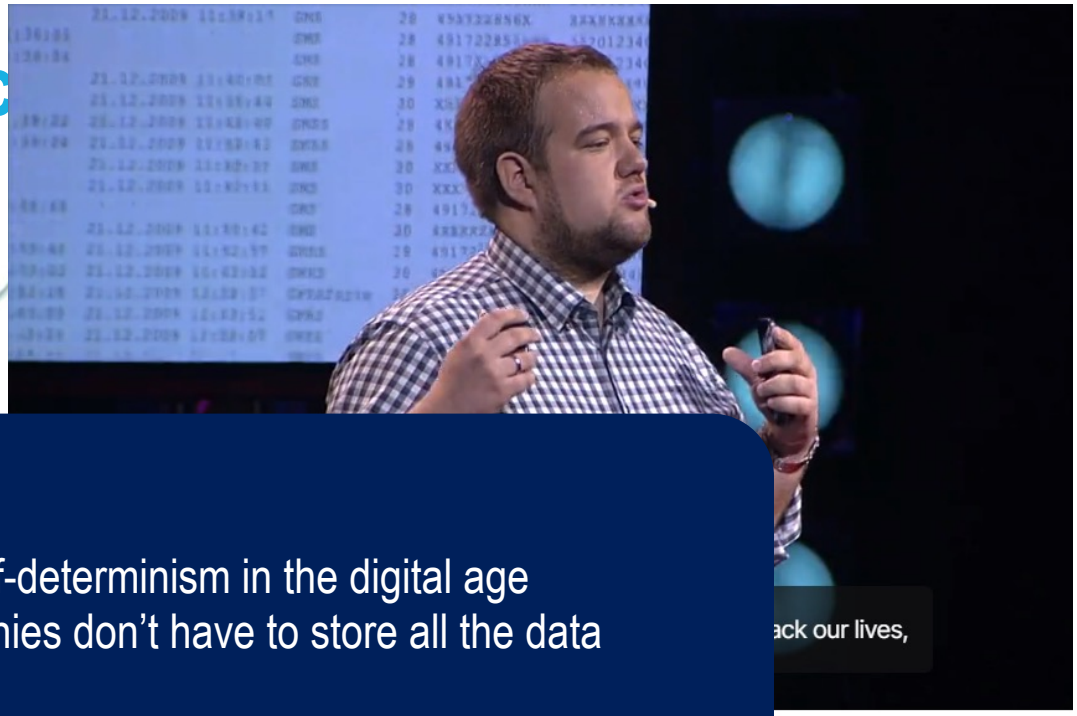
Project Profiler-Teams

Wir haben einen Mobilfunkdatensatz einer Person erhalten. Dieser erstreckt sich über den Zeitraum eines halben Jahres. Eure Aufgabe ist es in 3er Gruppen ein Profil dieser Person zu erstellen! Eurer Kreativität sind dabei keine Grenzen gesetzt!

Aufgabe: Das Profil der Person soll von Ihnen als Plakat angefertigt werden. Am Ende des Profiler-Projects müssen Sie ihr erstelltes Profil der Person präsentieren.

Folgendes bevor Sie mit Hilfe von Processing das „profilen“ beginnen:

The ‘old’ teac



Self-determinism in the digital age
Companies don’t have to store all the data

Durchl

Im Folge

Speicher

Entscheide

"Was sich aus unseren Einkäufen
2014). http://crackedlabs.org/stu

ack our lives,

Like (52K)

Sh

Your phone company is watching

What kind of data is your cell phone company collecting? Malte Spitz wasn't too worried when he asked his operator in Germany to share information stored about him. Multiple unanswered requests and a lawsuit later, Spitz received 35,830 lines of code -- a detailed, nearly minute-by-minute account of half a year of his life.

Privacy & the Mobile Phone Network

~ understanding digital artifacts
~ learning computer science (here e.g. informatics systems, protocols)
~ include societal impact (here e.g. privacy, data retention)

? Focus on threats in the old teaching unit & implicitly maybe teaching students how to behave instead of giving them tool to develop an informed opinion?

-> Does a data driven / date science education perspective change something?

New accentuation of the teaching unit by aiming fostering Data Awareness

New objective:

Raise awareness and sensitivity to the collection and processing of location data in example of cellular network and additional contexts with collection and processing of location data in students' everyday lives

Teaching unit:
Data Awareness by exploring location data from cellular network
(Computing Classes in grade 5/6)

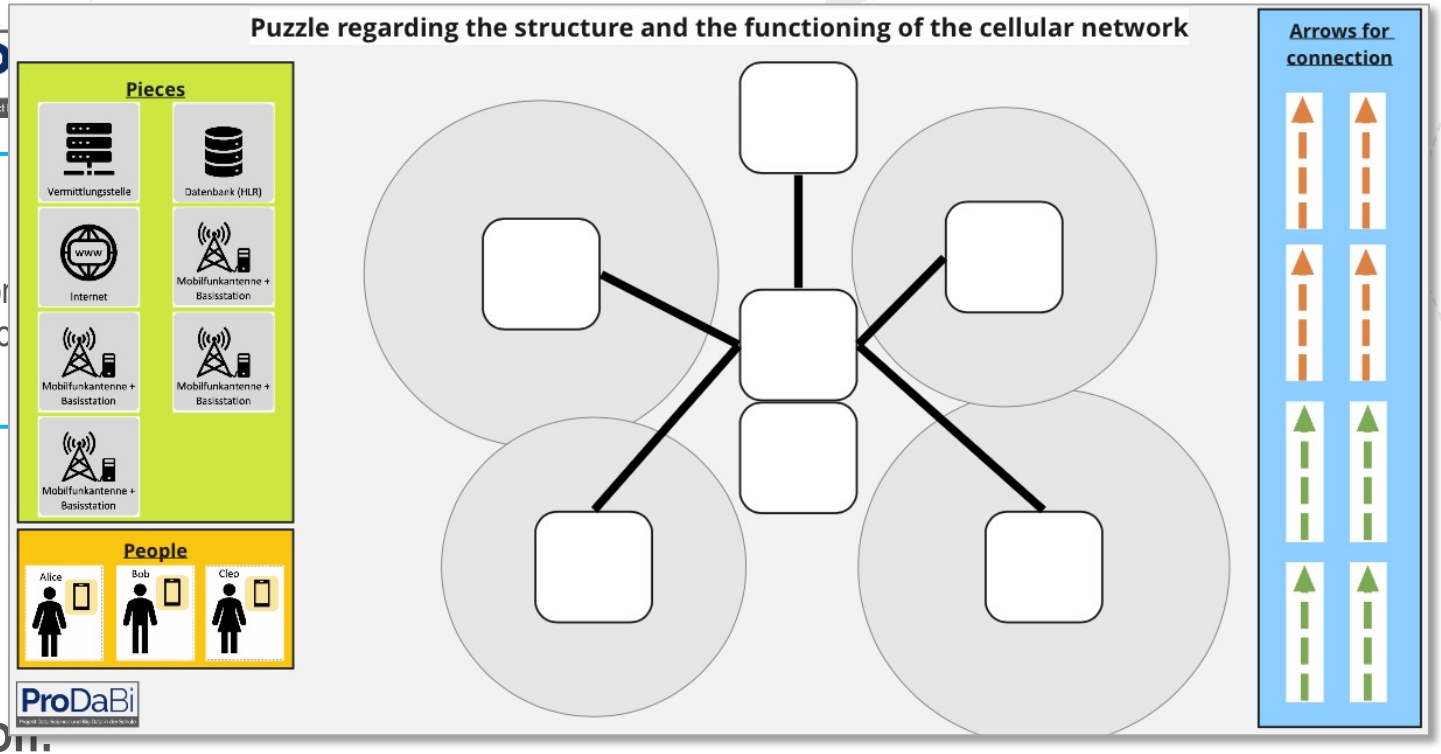
1st part:
Structure and functioning of the cellular network

2nd part:
exploration of given location data

3rd part:
additional contexts with collection and processing of location data

PADERBORN UNIVERSITY

1st part:
Structure and functio
the cellular netwo



Example for interaction:

Leading question:

How does the cellular network work and what data is explicitly and implicitly collected during interaction with the cellular network (e.g., making phone calls)?

Activities:

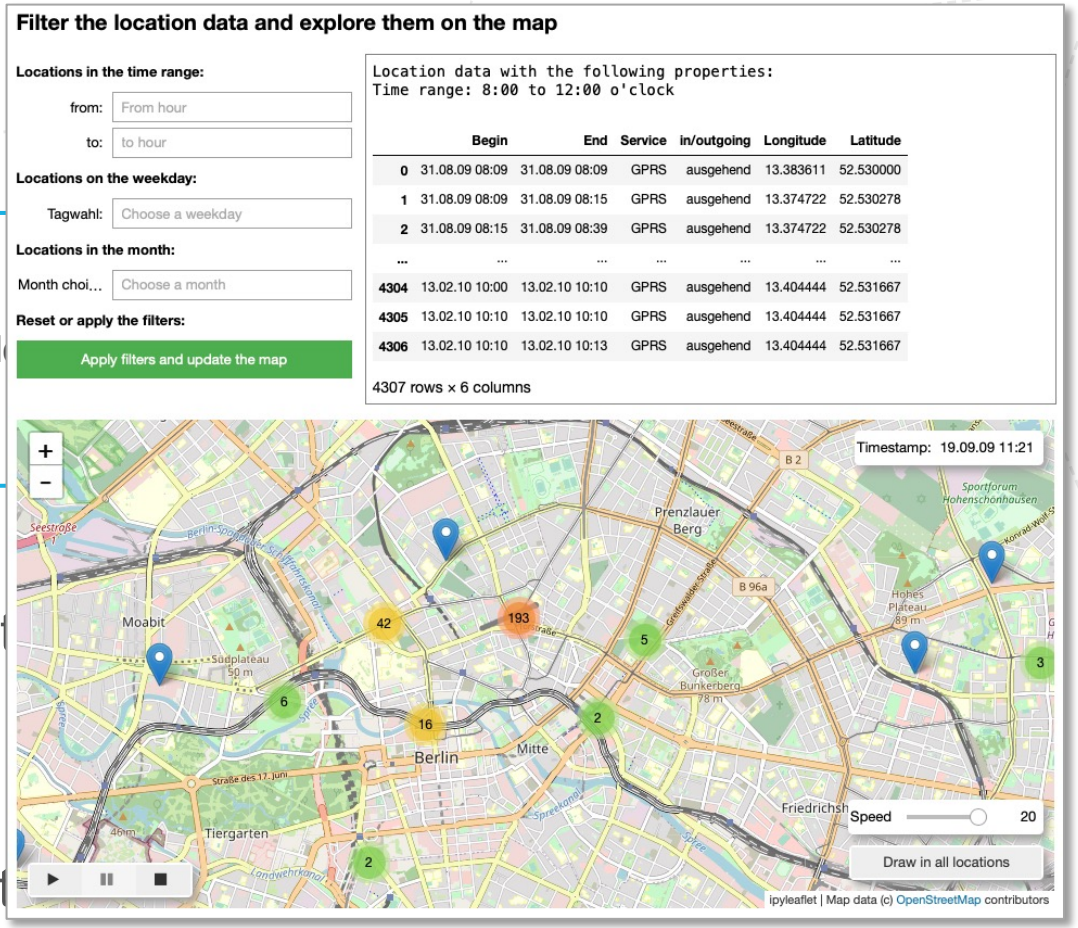
1. Thought experiment for introduction in the interaction context + explanation video
2. Puzzle as a model of a cellular network and simulation of a phone call
3. Reconstruction of the data collection regarding the primary purpose of establish the connection between two cellphones

PADERBORN UNIVERSITY

ProDaBi
Projekt Data Science und Big Data in der Schule

1st part:
Structure and functioning of the cellular network

expl

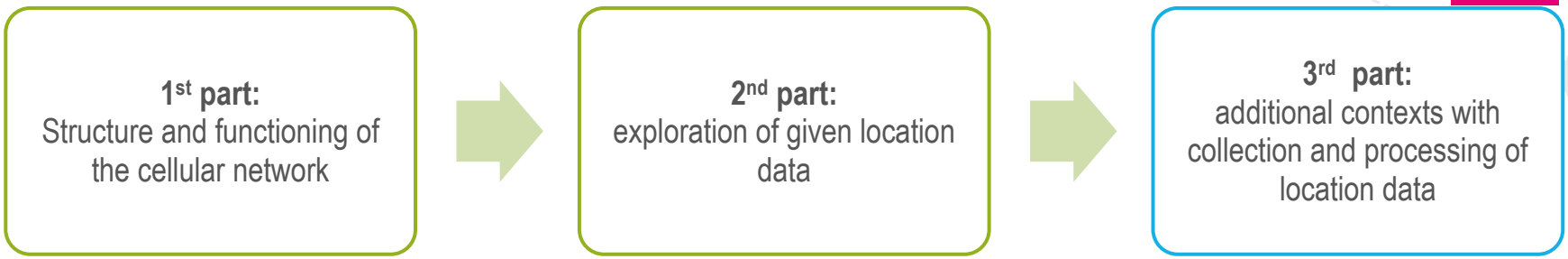


Leading question:

Which insights about a person can get

Activities:

1. Introduction in the given location
   location data from the cellular net
2. Introduction in a process of exploring the data with the provided web-application
   (1. ask a question, 2. filter the location data and visualize them on a map, 3. exploring and interpreting the location data, 4. answer the question/repeat the steps)
3. Exploration of the location data within a web-application and create a profile of the person



Leading question:

In which other contexts in our (students) everyday lives is location data collected and for which purposes is it processed/used?

Activities:

1. Application of the insights about the collection and processing of location data to other examples from students' everyday lives
2. Assess the data collection and processing regarding this examples

➔ Applying the learnings regarding Data Awareness to additional examples from their everyday lives

Overview of this teaching unit

1st part:
Structure and functioning of the cellular network

2nd part:
exploration of given location data

3rd part:
additional contexts with collection and processing of location data

- **Interaction between human and data-driven digital artifacts from students' everyday lives**
- **Explicit and implicit data collection during this interaction examined**
- **Primary and secondary purposes for using and processing the collected data examined and evaluated**
- **Application of students' learnings to other contexts from students' everyday lives where location data is collected**

New accentuation of the teaching unit by aiming fostering Data Awareness

<u>New objective:</u> Raise awareness and sensitivity to the collection and processing of location data in example of cellular network and additional contexts with collection and processing of location data

<u>So what is new?</u>

- **Data collection should be viewed in a more differentiated way and more focused** (focus at the beginning is now no longer the architecture of the cellular network)
- **Data processing has also primary purposes that is useful and good, so there are also good purposes for processing and using collected personal data** (besides sometimes negatively connoted secondary purposes)
- **Application of the learnings to other contexts from students' everyday lives at the end**
- **Data Awareness also gives new terms that help as "tools" to describe, understand and evaluate such examples**

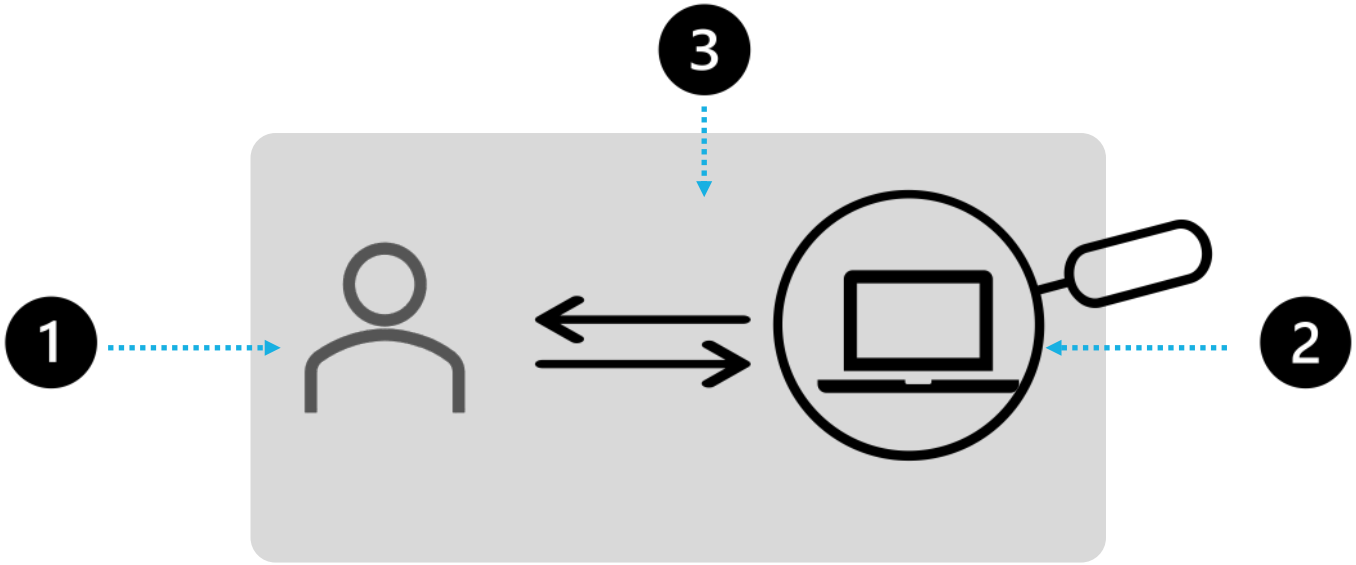
PADERBORN UNIVERSITY

ProDaBi
Projekt Data Science und Big Data in der Schule

Deutsche Telekom Stiftung

Interactionfocused view on computing education

Hybrid Interaction System



1. The role of the human: “program or be programmed”
2. The role of the artifact: replacement, augmentation, symbiosis
3. The role of the hybrid system: shaping and being shaped

SCHULTE, Carsten and BUDDE, Lea, 2018. A Framework for Computing Education: Hybrid Interaction System: The need for a bigger picture in computing education. In: *18th Koli Calling International Conference on Computing Education Research (Koli Calling ’18)*. Koli, Finland: ACM. 22 November 2018

Rationale

Machine Behaviour, 2019

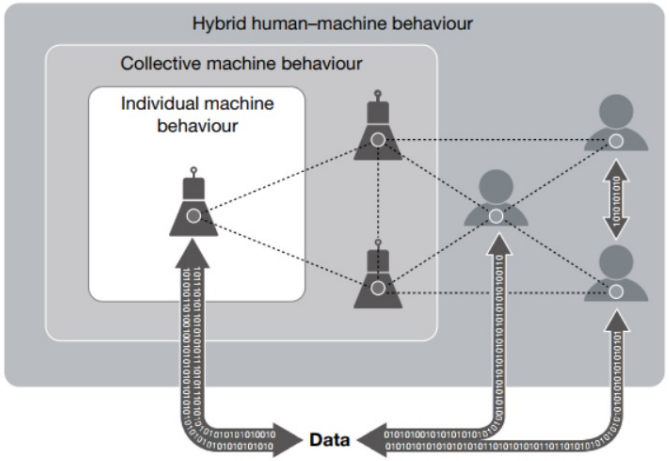


Fig. 4 | Scale of inquiry in the machine behaviour ecosystem. AI systems represent the amalgamation of humans, data and algorithms. Each of these domains influences the other in both well-understood and unknown ways. Data—filtered through algorithms created by humans—influences

Rahwan, Iyad ; Cebrian, Manuel ; Obradovich, Nick ; Bongard, Josh ; Bonnefon, Jean-François ; Breazeal, Cynthia ; Crandall, Jacob W. ; Christakis, Nicholas A. ; u. a.: Machine behaviour. In: Nature Bd. 568 (2019), Nr. 7753, S. 477–486. — tex.ids= Rahwan.2019, rahwanMachineBehaviour2019a

Computational Thinking, 2006

"Ideas, not artifacts"

1970 1980 1990 2000 2010 2020

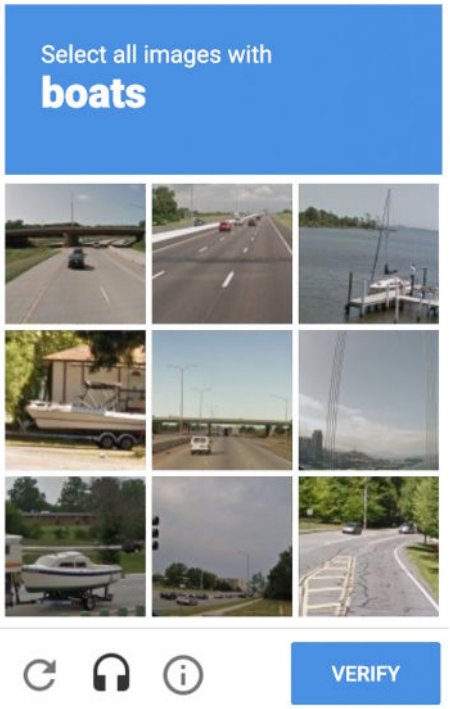
WING, Jeanette M, 2006. Computational Thinking. *Communications of ACM*. March 2006. Vol. 49, p. 33–35. DOI 10.1145/1118178.1118215.
p. 35

Framework of Data Awareness

1. Introducing data-driven digital artifact and the interaction between human and data-driven digital artifact
2. Data-driven version of Hybrid Interaction System for the Framework of Data Awareness
3. Facets for fostering Data Awareness
4. Another example for a teaching unit on Data Awareness

Example for data-driven digital artifacts

Google once provided the US Pentagon with such ML models for military drones

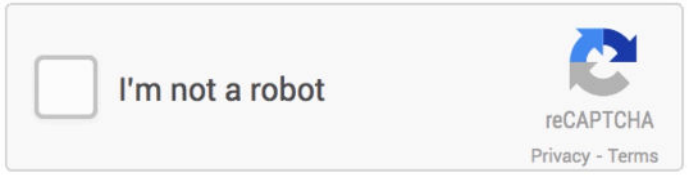


Source: Google reCAPTCHA

Which data is collected?
The images selected by the user

For what purpose these data is collected?
Distinguish between human and technology and also train a ML model for image recognition

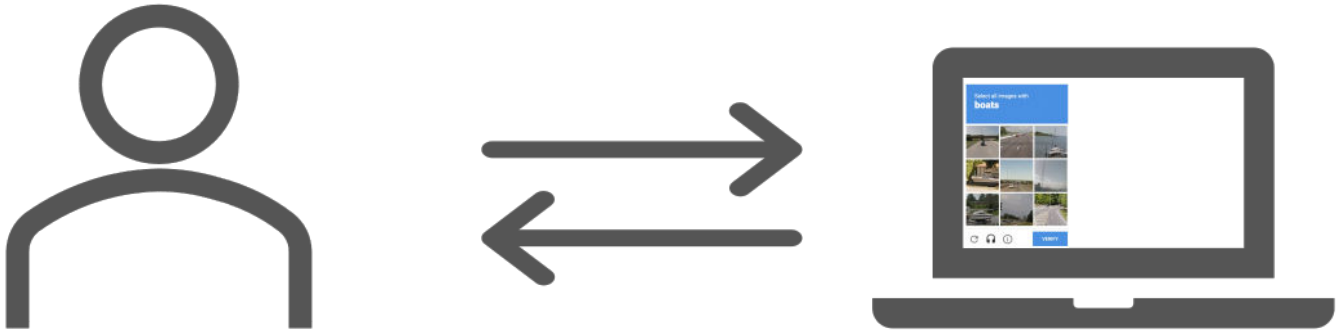
How does it look here?
Characteristic mouse movement of the user is tracked for distinction



Source: Google reCAPTCHA
developers.google.com/recaptcha

What happens during interaction with data-driven digital artifacts?

Interaction process



user

data-driven digital artifact

Collection and processing of personal data

Are school students already aware of this data collection and processing? (related empirical studies)

Data collection:

No awareness and understanding ⟷ Good awareness and understanding

(Pangrazio & Selwyn, 2020): don‘t know what personal data are and how it is collected

(Tedre et al., 2020): no adequate understanding

(Bowler, 2017): know that data is collected

(Bucher, 2017): some social media users know that data is collected

(Keen, 2020): some of them had ideas; some also about tracking

(Gabriele & Chiasson, 2020): only about fitness tracker

Seems to be not so good.

Processing and using collected data:

No awareness and understanding ⟷ Good awareness and understanding

(Bowler, 2017), (Keen, 2020), (Gabriele & Chiasson, 2020): no ideas what happens after collection

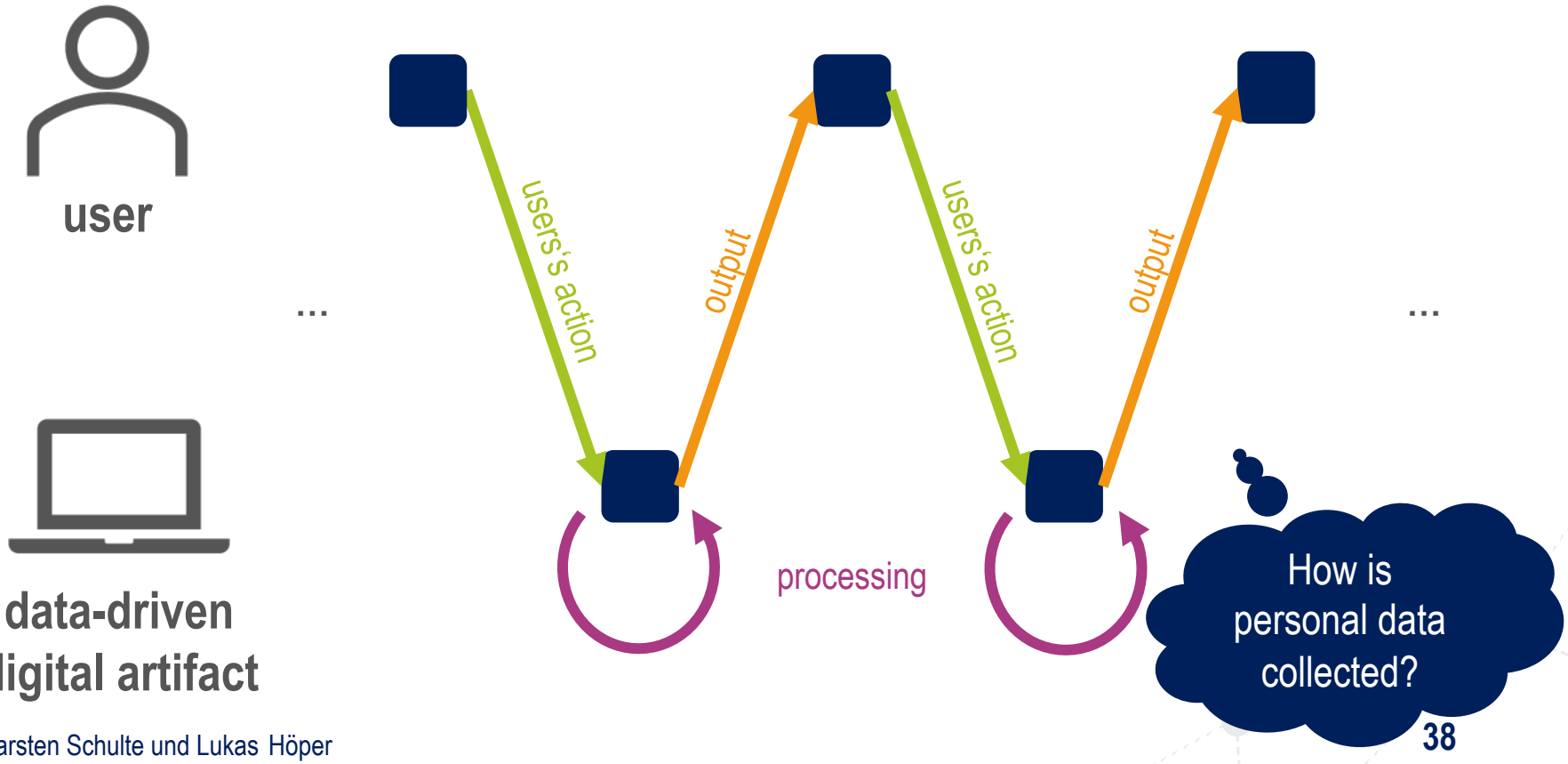
(Eslami et al, 2015): 62% were not aware about usage of curation algorithms

(Tedre et al., 2020): no adequate understanding

(Pangrazio & Selwyn, 2020): couldn’t describe implications

Seems to be not good.

Interaction between human and data-driven digital artifact

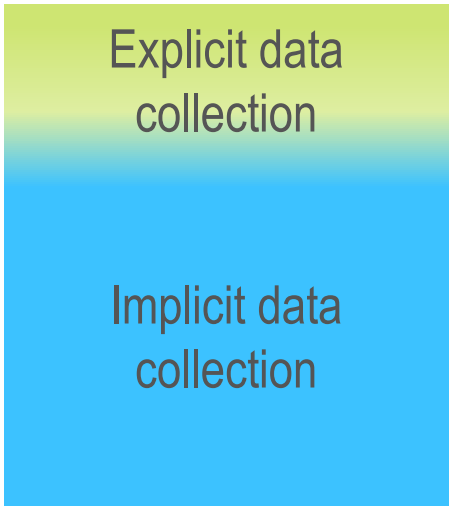


Collection of (personal) data

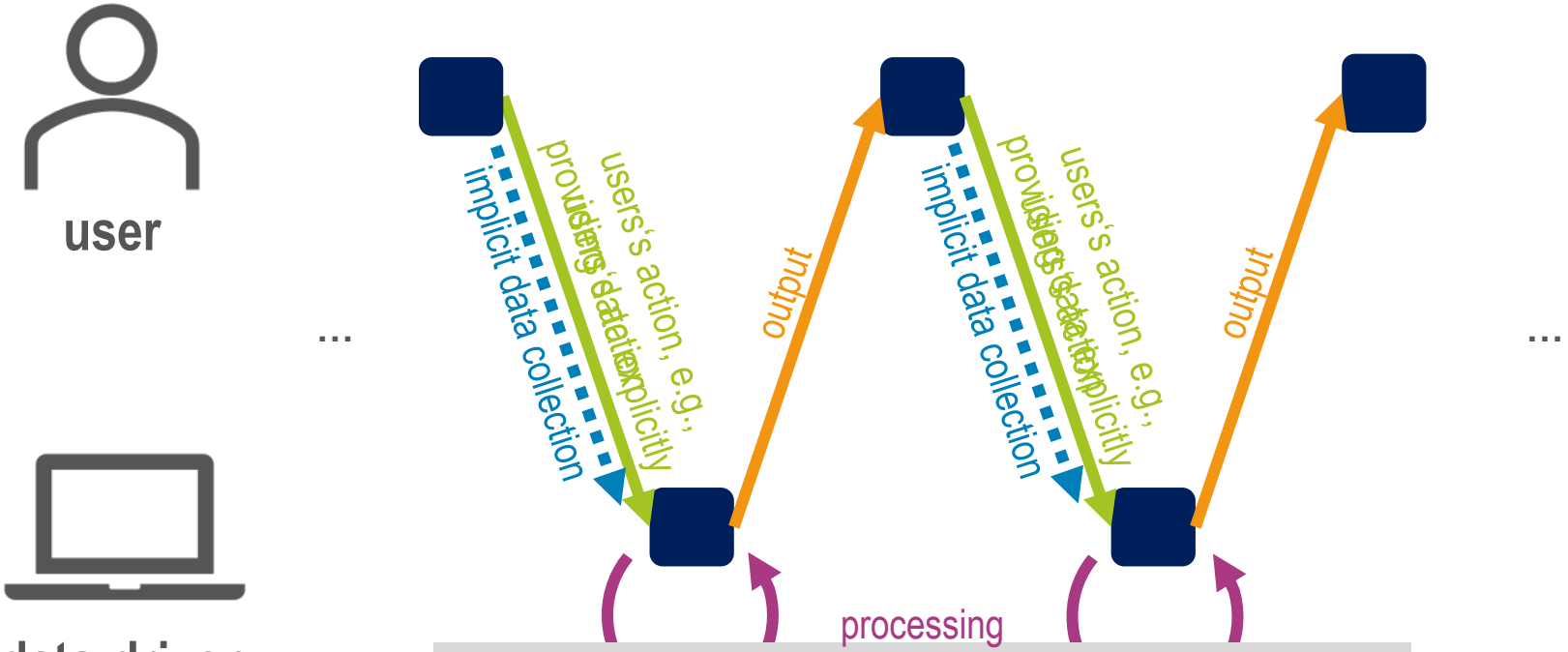
Taxonomy of collected data:

- **provided data** is data which is created actively by the user
- **observed data** is data which is gathered by observation and recording, and of which the user is not necessarily aware
- **derived data** is data which is generated by processing of existing data
- **inferred data** is data which is generated by probability-based analytically processing

(OECD, 2014)



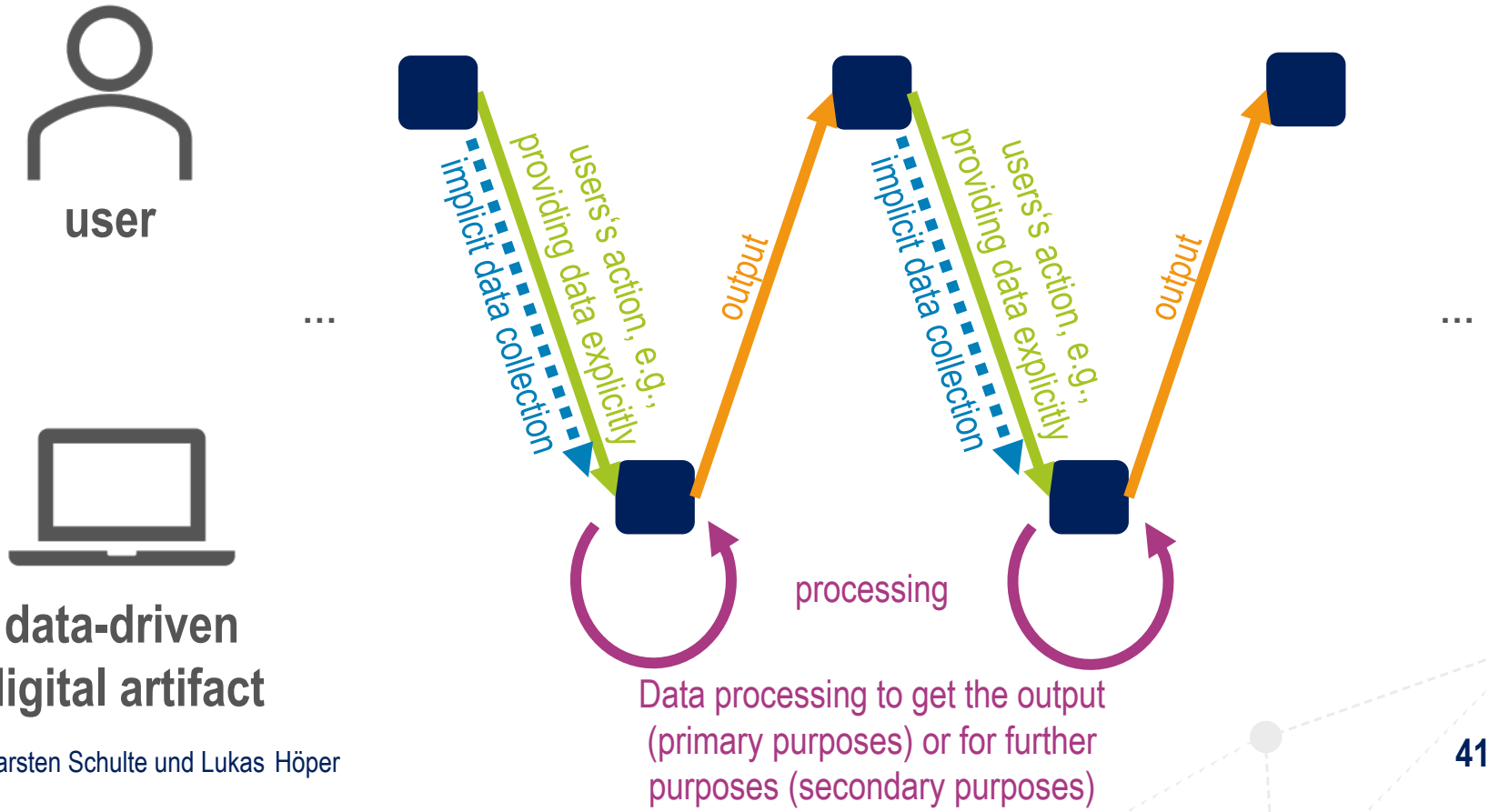
Interaction between human and data-driven digital artifact



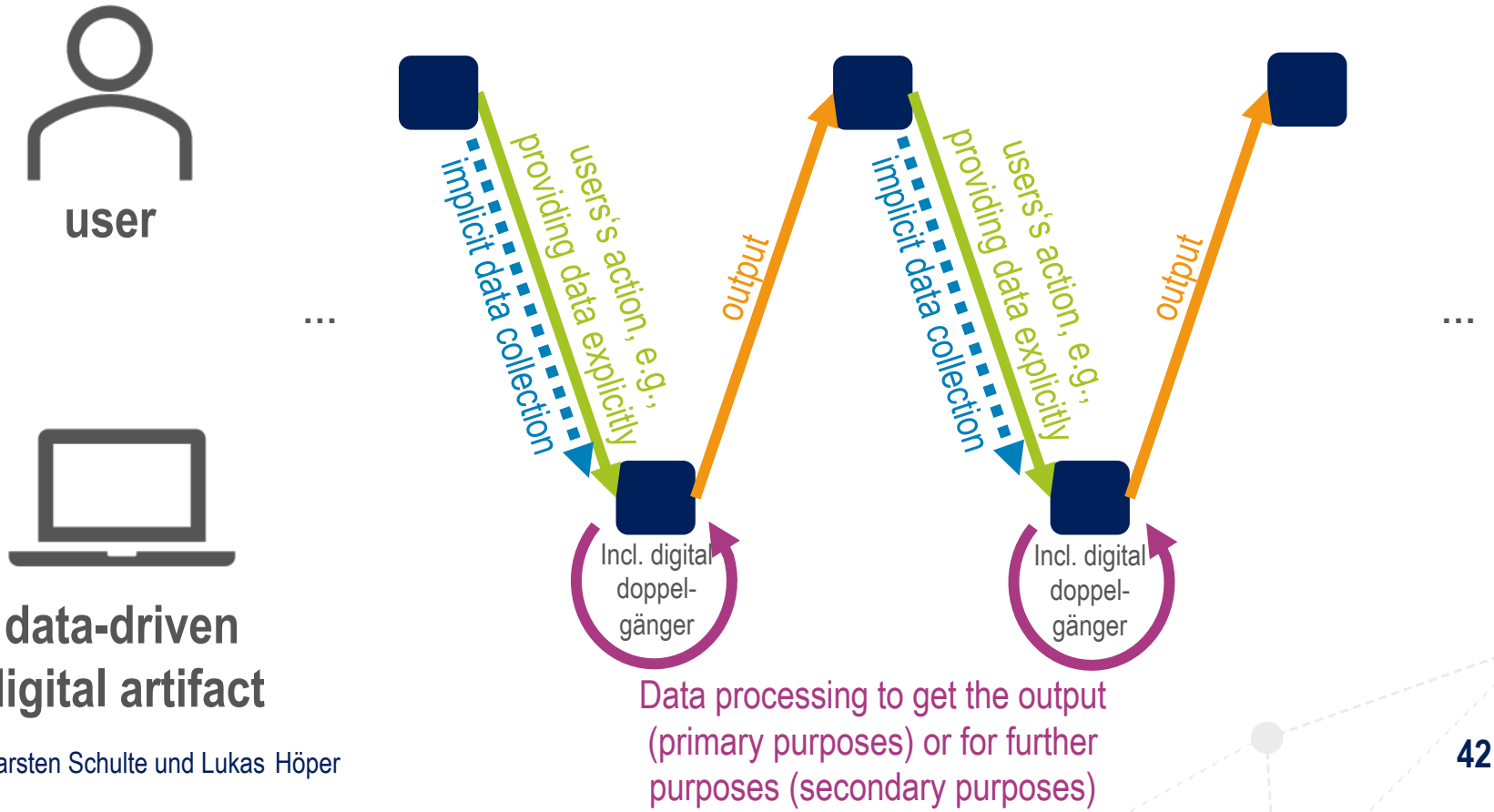
Explicit data collection: User's action during the interaction aims to create this data (e.g., posts of texts or images, search terms, …).

Implicit data collection: User's action during the interaction aims to do something (not create data) where data are collected incidentally.

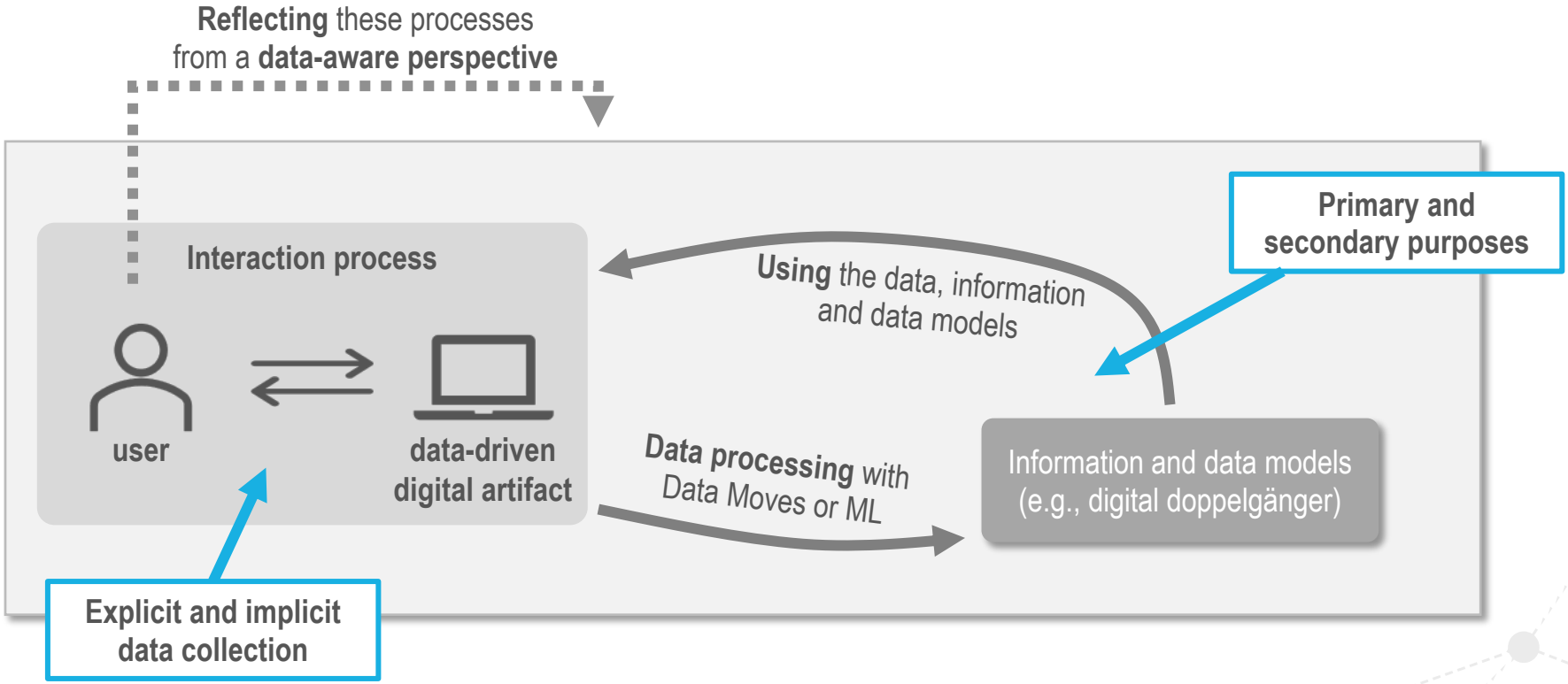
Interaction between human and data-driven digital artifact



Interaction between human and data-driven digital artifact



Data-driven version of Hybrid Interaction System as a model for the Framework of Data Awareness



Data awareness is the awareness of the explicit and implicit collection of personal data and its processing and use for primary and secondary purposes during interactions with a data-driven digital artifact.

Framework of Data Awareness

Data awareness is the awareness of the explicit and implicit collection of personal data and its processing and use for primary and secondary purposes during interactions with a data-driven digital artifact.

Facets for fostering Data Awareness:

1. Exemplary context of interaction between human and data-driven digital artifact
2. Explicit and implicit data collection during this interaction
3. Primary and secondary purposes for using and processing the collected data
4. (If applicable) constructing a profile or digital doppelgänger as a data model

Steps of the teaching unit

Data Awareness by exploring location data from cellular network

1. Introduction into the interaction context
2. Explicit and implicit data collection
3. Primary Purpose for using and processing the data
4. (if applicable: profile)
5. Secondary purpose for using and processing the data
6. Application to other contexts

1st Part: structure and functioning of the cellular network

2nd Part: exploration of given location data

3rd Part: additional contexts with collection and processing of location data

Example for grades 5 and 6

Another example for grades 8-10

Teaching unit 2:
Data Awareness by exploring recommender systems
(Computing Classes in grades 8-10)

1st part: Movie recommendations and data collection by a streaming service

2nd part: structure and functioning of a movie recommender system

3rd part: secondary use of collected data by a streaming service

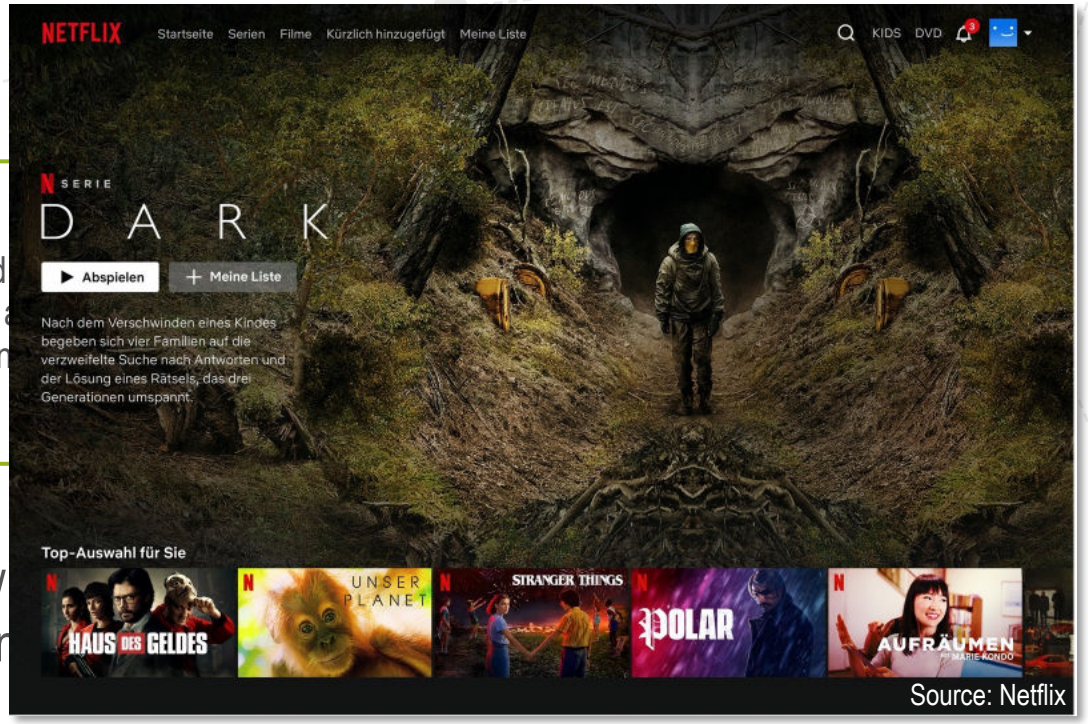
4th part: additional contexts with recommender systems

PADERBORN UNIVERSITY

ProDaBi
Projekt Data Science und Big Data in der Schule

1st part:
Movie recommendations and data collection by a streaming service

2nd part:
structure and functioning of a movie recommen system



Source: Netflix

Leading question:

What are movie recommendations, w
helpful and which data can a streamin

Activities:

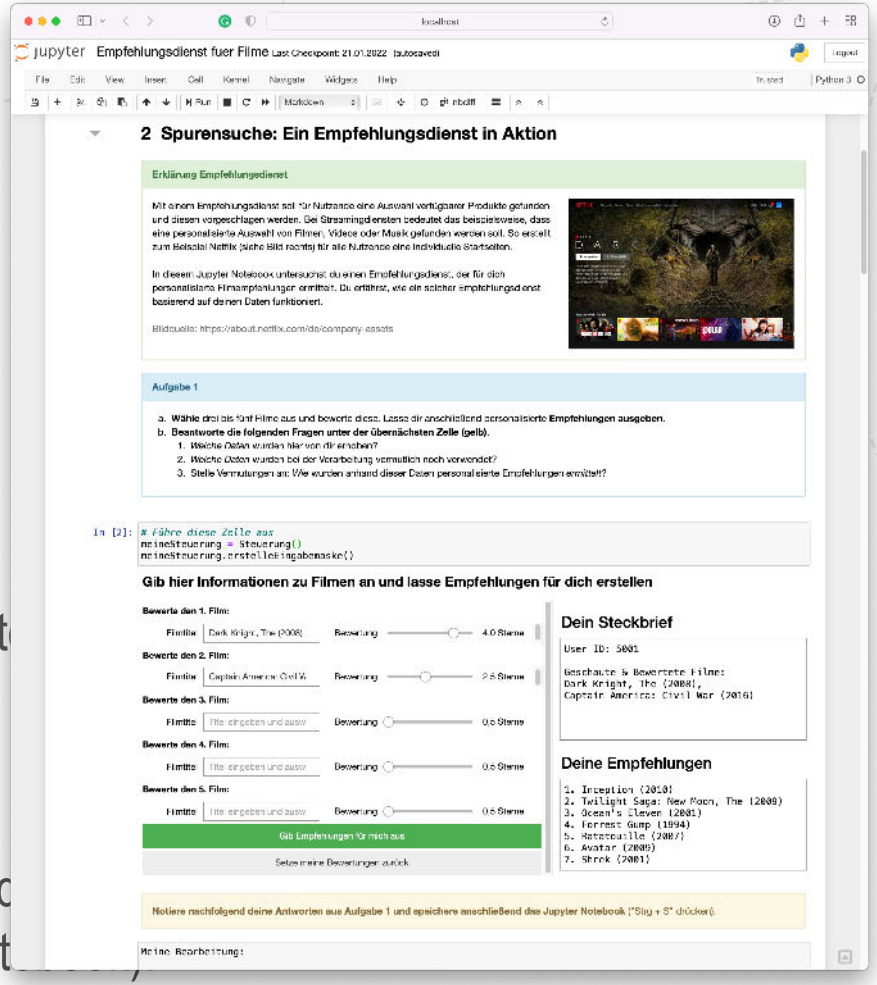
1. Problematisation with the example of the personalised Netflix start page
2. Recommendation game: Developing the meaning of "movie recommendations" and identifying helpful information for this purpose.
3. Interaction with a provided recommender system and reconstruction of explicit and implicit data collection (within a Jupyter Notebook)

PADERBORN UNIVERSITY

ProDaBi
Projekt Data Science und Big Data in der Schule

1st part:
Movie recommendations and data collection by a streaming service

2nd part:
structure and functioning of a movie recommender system

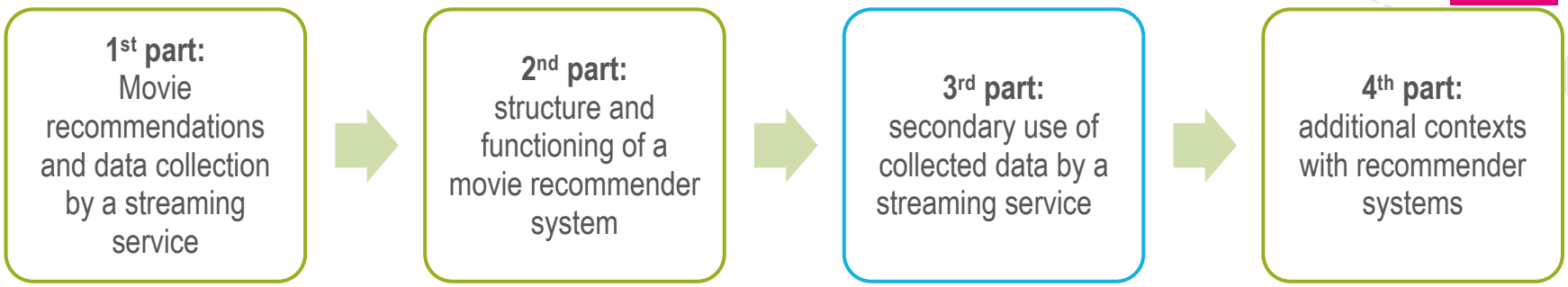


Leading question:

How can rating data (and others) be used to aut
recommendations?

Activities:

1. Reconstruction of a exemplary process of c
   (step-by-step within a provided Jupyter Not
   - personalised recommendations for the students on the basis of the students' own rating data
   - oriented to the k-nearest-neighbour method
   - Explicit and implicit "rating" data used

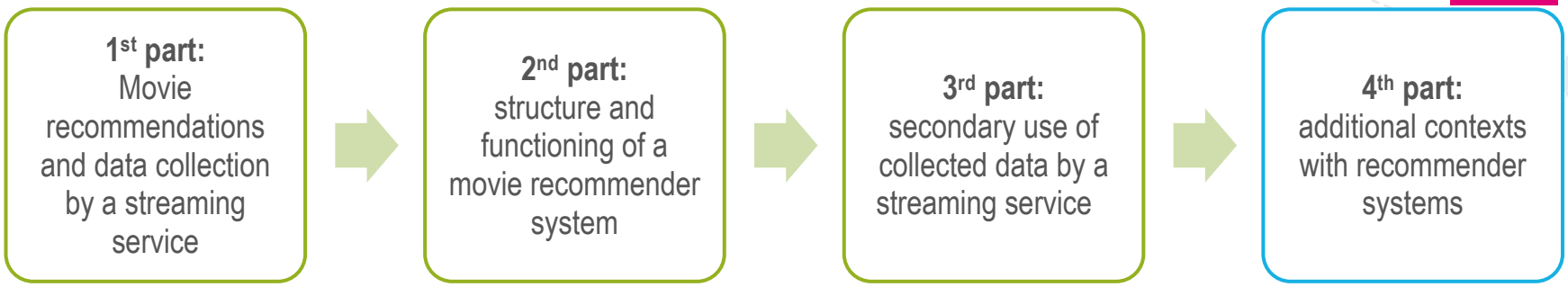


Leading question:

What else could the personal data be used for in addition to the purpose of identifying personalised movie recommendations?

Activities:

1. Role play on a fictional secondary use (secondary purpose) of the collected data in an exemplary recommendation service: A proposal to improve the profit of a streaming service is discussed in a head of department meeting ("personalised pay barrier").
2. Evaluation of the role play and assessment of the interaction system



Leading question:

In what other contexts are recommender systems used and what data are collected there and processed for what purposes?

Activities:

1. Reconstruction of interaction contexts from the students' everyday lives in which recommender systems are used with regard to the facets of data awareness.
2. Presentations of the findings and evaluation of the data collection and processing in the different contexts as well as in general

➔ Students apply their learnings to other contexts where recommender systems are used

PADERBORN UNIVERSITY

ProDaBi
Projekt Data Science und Big Data in der Schule

Deutsche Telekom Stiftung

Overview of the teaching units on Data Awareness

1. Introduction into the interaction context
2. Explicit and implicit data collection
3. Primary Purpose for using and processing the data
4. (if applicable: profile)
5. Secondary purpose for using and processing the data
6. Application to other contexts

Data Awareness by exploring location data from cellular network	Data Awareness by exploring recommender systems
1st Part: structure and functioning of the cellular network	**1st Part:** movie recommendations and data collection by a streaming service
	2nd Part: structure and functioning of a movie recommender system
2nd Part: exploration of given location data	**3rd Part:** secondary use of collected data by a streaming service
3rd Part: additional contexts with collection and processing of location data	**4th Part:** additional contexts with recommender systems

PADERBORN UNIVERSITY

ProDaBi
Projekt Data Science und Big Data in der Schule

Deutsche Telekom Stiftung

Thank you very much for your attention!

www.prodabi.de/en

Literature

- Taina Bucher. 2017. The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms. Information, Communication & Society 20, 1 (2017), 30–44. https://doi.org/10.1080/1369118X.2016.1154086
- Leanne Bowler, Amelia Acker, Wei Jeng, and Yu Chi. 2017. “It Lives All around Us”: Aspects of Data Literacy in Teen’s Lives. Proceedings of the Association for Information Science and Technology 54,1 (2017), 27–35. https://doi.org/10/ghsdj9
- Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I Always Assumed That I Wasn’t Really That Close to [Her]": Reasoning about Invisible Algorithms in News Feeds. In Proceedings of the33rdAnnualACMConferenceonHumanFactorsinComputingSystems.ACM,SeoulRepublicofKorea,153–162. https://doi.org/10/gfgxg9
- Sandra Gabriele and Sonia Chiasson. 2020. Understanding Fitness Tracker Users’ Security and Privacy Knowledge, Attitudes and Behaviours. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. ACM, Honolulu HI USA, 1–12. https://doi.org/10/gnqv9n
- Caroline Keen. 2020. Apathy, Convenience or Irrelevance? Identifying Conceptual Barriers to Safeguarding Children’s Data Privacy. New Media & Society 24, 1 (2020). https://doi.org/10/gndjz8
- OECD. 2014. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. https://doi.org/10.1787/9789264196391-en
- Luci Pangrazio and Neil Selwyn. 2020. Towards a School-Based ‘Critical Data Education’. Pedagogy, Culture & Society 29, 3 (2020), 431–448. https://doi.org/10/gj5kc9
- Matti Tedre, Henriikka Vartiainen, Juho Kahila, Tapani Toivonen, Ilkka Jormanainen, and Teemu Valtonen. 2020. Machine Learning Introduces New Perspectives to Data Agency in K—12 Computing Education. In 2020 IEEE Frontiers in Education Conference (FIE). IEEE, Uppsala, 1–8. https://doi.org/10/ghrk22