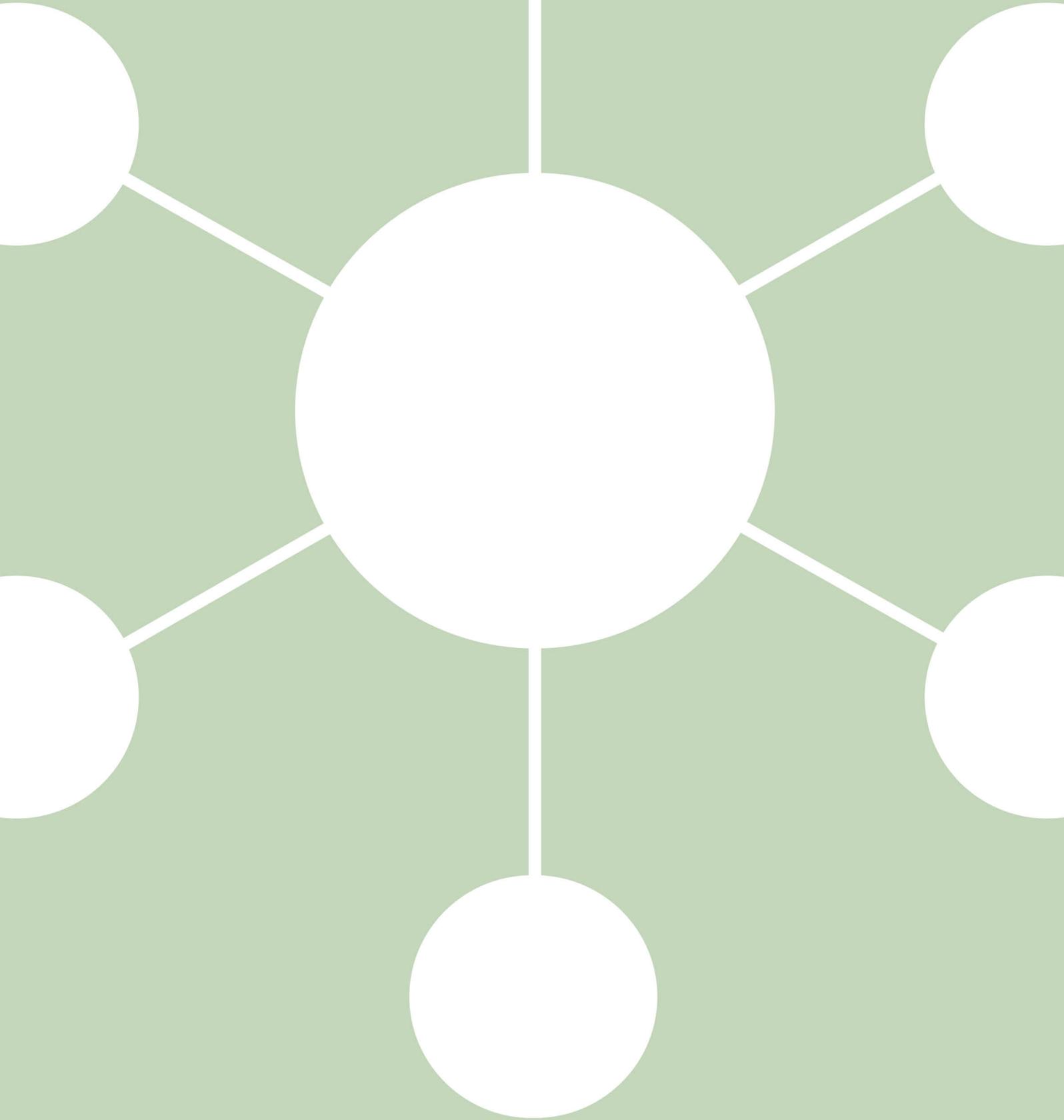


**WHERE DOES THE
DATA GO?**



Data collection → Education

How is my data collected online and where does it go? Answering these questions is not as simple as it may seem, as data collection online often takes place in opaque ways. Apart from actively disclosing our personal data, also our actions online are constantly tracked, for example through cookies, leading to a digital fingerprint for every user. A large part of the traces we leave online consist of so-called metadata. Although they are often downplayed, metadata include a great deal of personal information about us: for example, our location, how much time we spent on certain websites or apps, how often and how long we contacted whom, and how much time we spent reading which page of our e-book.

All these data are initially collected and stored by the platforms, which then often sell them to so-called data brokers that combine these datasets with data from other sources. Using their vast datasets, data brokers can, among others, create detailed psychological profiles, which are then sold to advertisers. All of this is happening because a lot of money can be made from data. However, data never portray reality in a neutral and objective manner. Instead, they are massively influenced by their context of collection, and conversely create their own reality. Even with stricter data collection laws such as the General Data Protection Regulation (GDPR), many forms of data collection and analysis remain possible (such as the use of metadata) and sometimes only gain a legal framework through these laws. Moreover, personal data are often only pseudonymised – a process that comes with considerable risk – instead of following the principle of data minimisation and collecting as little data as possible.

Also the field of education is increasingly being pervaded with powerful data infrastructures. Nearly every digital tool that is used in educational contexts collects data, as especially interactive systems only work if they know their users. Thus, the suppliers – often private, commercial companies – collect data from millions of learners, for example about exam results and teaching content, but also about social behaviour, classroom climate, absences, facial expressions, posture, or tone of voice. These data are analysed by the companies' algorithms (so-called 'Learning Analytics'), which then – sometimes in real-time – create personalised recommendations. The more data is collected, the better these predictive systems work. But also outside the classroom, data is being collected. Publishers of digital school books, for example, can use platforms such as 'scook', which allow them to see which chapters are most used by teachers. Also agencies that are monitoring schools can get access to some data from school administration software. Apart from risks of surveillance through the massive data collection of digital educational tools, problems lie in the decreasing privacy of students; the opaque nature of the data collection and analysis; the sometimes life-long storage of educational data; and the monopolistic structures of US software companies that are often behind digital learning software, and that come with massive risks for data protection. The question of data protection becomes even more controversial when data of minors is involved. Furthermore, commercial companies gain a more and more influential role in the education sector, and increasingly also computer scientists are able to influence pedagogical decisions. Moreover, educational data constitute a lucrative business: on the basis of the data they collect, these companies develop new theories of learning and new software models – a process which takes place covertly as these systems are protected by patent and copyright law.

Data collection by the digital board “Padlet”

The popular tool “Padlet”, a web-based digital board that can be used to upload tasks and texts, constitutes a succinct example of data collection in the education sector. A study that recently examined the tool found that, among others, data about the user's browser, location and visited websites were shared with numerous tracking and marketing services. Among others, these included Google Analytics, Alexa Metrics and Quantserve (see fig. 1). The companies behind these services, such as Quantcast, create detailed profiles based on their data (which is sometimes collected from over 100 million different sources), that are used to make predictions about user behaviour.

→ Questions

- Which (meta)data are collected openly as well as without the users' knowledge?
- Are all of these data points necessary? For what purposes? Is the principle of data minimisation adhered to?
- Which advantages and disadvantages do other learning environments – without the use of data collection – have?
- Where and for how long are the data kept? Are they reused in the next school year or even for other purposes?
- If a commercial company is behind the tool: How does the company earn money?
- Do the students have the right and the possibility to see the data that is stored about them?
- Who is behind the system? Where is the supplier based and how does this affect applicable laws for data protection?

➔➔ Seeing the signs

- Does the supplier promise personalised learning?
- Are individual profiles created for minors?
- Is the company from the US?
- Are data being pseudonymised?
- Is the supplier using jargon?
- Are you uncertain which data are being collected or about the pedagogical use of the collection of some of this data?

➔➔ Then you should definitely pay attention and make further inquiries!

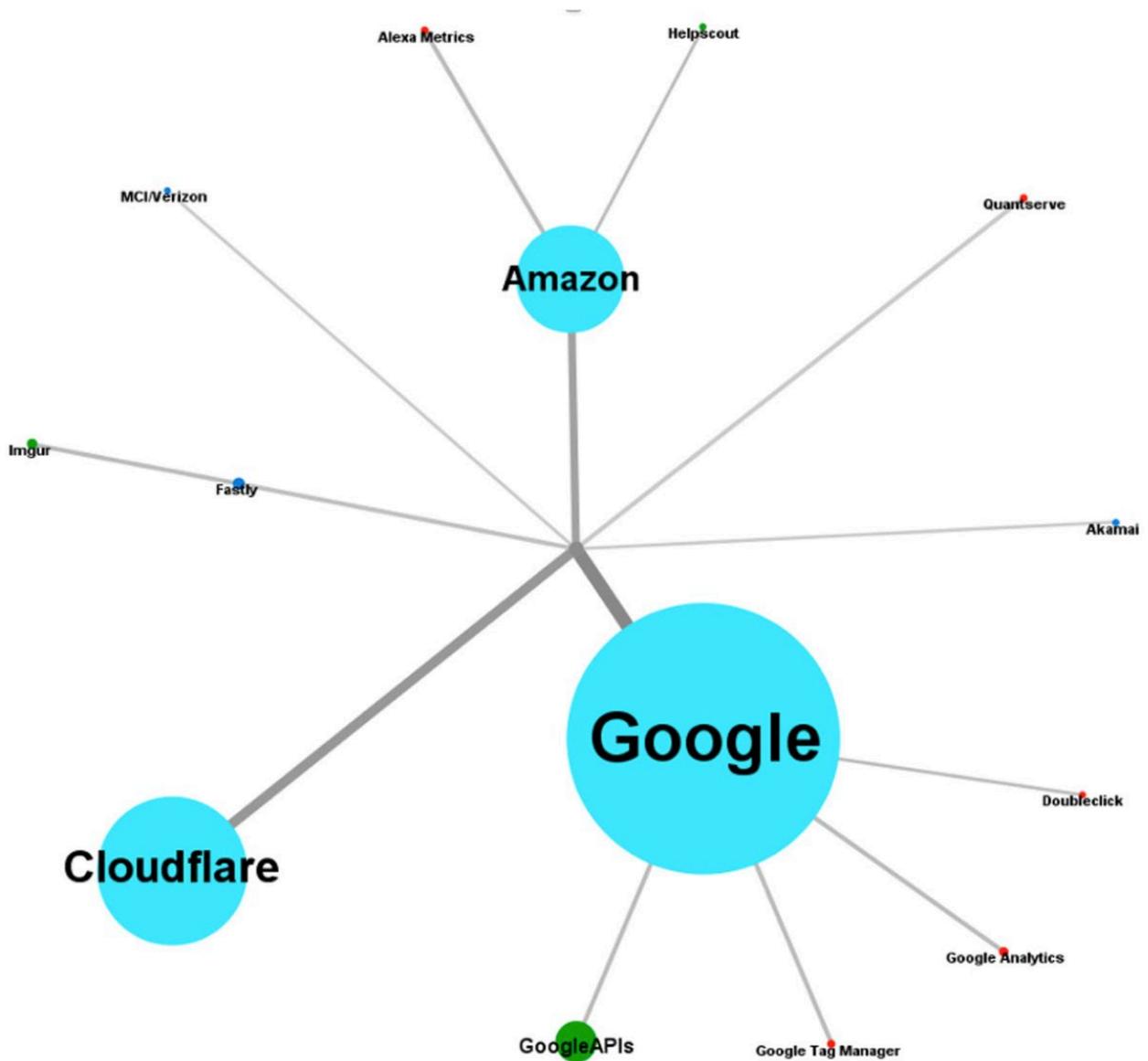


Fig.1: Data traffic during the use of the digital tool "Padlet";
Turquoise = Server host in the US;
Blue = Server host in the EU;
Green = Functional service;
Red = Tracking and marketing service.

Source: Jornitz, S. and Macgilchrist, F. (2020). Optimierung outgesourced? Die Transformation pädagogischer Sprachspiele und Praxisformen durch Datafizierung. DGfE, Cologne, March 2020.

Literature and References

Hartong, S. (2019). Learning Analytics und Big Data in der Bildung. Gewerkschaft Erziehung und Wissenschaft. <https://www.gew.de/index.php?eID=dumpFile&t=f&f=91791&token=702ec8d5f9770206a4aa8a1079750ec9021b90bf&sdownload=&n=Learning-analytics-2019-web-IVZ.pdf>

Hartong, S. (2020). Algorithmisierung von Bildung - Über schrumpfende Spielräume für demokratisches (Ver-) Handeln und warum die EdTech-Industrie nicht das einzige Problem ist. <https://denk-doch-mal.de/wp/sigrid-hartong-algorithmisierung-von-bildung/>

Sander, I. (2020). Warum unsere Daten wertvoll sind. <https://www.politische-bildung.nrw.de/digitale-medien/digitale-demokratiekompetenz/big-data/>

Watters, A. (2020). School Work and Surveillance. Hack Education 30 April. <http://hackededucation.com/2020/04/30/surveillance>

Williamson, B. (2016a). Critical questions for big data in education. code acts in education 2 June. <https://codeactsineducation.wordpress.com/2016/06/02/critical-questions-for-big-data-in-education/>

Williamson, B. (2016b). Powerful algorithms in education. Code Acts in Education research summary 1. University of Stirling. <https://codeactsineducation.files.wordpress.com/2016/03/powerful-algorithms-in-education-code-acts-summary-1.pdf>

Williamson, B. (2016c). Digital data in education policy. Code Acts in Education research summary 2. University of Stirling. <https://codeactsineducation.files.wordpress.com/2016/03/digital-data-in-education-policy-code-acts-summary-2.pdf>

Williamson, B. (2016d). High-tech theories of human learning. Code Acts in Education research summary 4. University of Stirling. <https://codeactsineducation.files.wordpress.com/2016/03/high-tech-theories-of-human-learning-code-acts-summary-4.pdf>

This work is licensed under a Creative Commons Attribution-No Derivatives 4.0 International License (CC BY-ND 4.0).

Authors: Ina Sander, Heidrun Allert, Karin Amos, Paula Bleckmann, Izabela Czarnojan, Annina Förschler, Sigrid Hartong, Sieglinde Jornitz, Manuel Reinhard.



UNBLACK THE BOX is a network initiative founded in 2019 by researchers from education science, sociology, information technology, media and health education, as well as teachers in schools, universities and pedagogical training. Our goal is to enable educational institutions and teachers to respond to the growing datafication and digitization of education with enlightened, critical and conscious decision-making, even without extensive IT knowledge.

<https://unblackthebox.org/unblack-the-box/>