Data are Capta — Data Diaries as Method for Exploring the Generated Nature of Data

ROSA VAN KONINGSBRUGGEN, Bauhaus-Universität Weimar, Germany
EVA HORNECKER, Bauhaus-Universität Weimar, Germany

To explore how we can change our understanding of data as something which is given and 'just appears', to something which is actively created and generated, and thus has subjective or deliberative elements, we describe our method of *Data Diaries*. We show examples of Data Diary creations from our students, created during two separate courses, and reflect on the developments we saw during theses assignments. Through the Data Diaries, we show how course participants' understandings of data changed towards embracing feminist perspectives and considering data as capta.

Additional Key Words and Phrases: Data Physicalisation, Sensification, Data Diaries, Notion of Data, Data Literacy, Teaching

ACM Reference Format:
Rosa van Koningsbruggen and Eva Hornecker. 2018. Data are Capta — Data Diaries as Method for Exploring the Generated Nature of Data. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Currently, data are associated with maths, numbers, diagrams, and graphs [4, 10], which creates the illusion that data are factual, objective, impersonal, and abstract [1, 4]. This is in stark contrast to how data are generated and how we encounter them in our lives [2]—which is messy [8, 13], local [11], and sometimes even emotional [6, 7]. The perceived factualness of data make them appear to be a given [3, 4], rather than something which has to be created [2]. To counter this, Drucker suggests using a different term: whereas the word data means 'given' (Latin, plural of *"that is given"*), the word 'capta' –meaning 'actively taken'– highlights that data are generated and need to be interpreted as data, in order for them to function as such [3, 4].

We have found that in teaching, *Data Diaries* [18] have been useful in getting students to realise that their factual view of data is biased towards 'given data' and that data are actually created, selected, and produced. Instead of picking an existing data set or data source, having to collect data from their own life engendered reflection on what and how they were collecting. Furthermore, creating representations that do not take the form of standard visualisations, and instead focus on materiality and touch, strengthened this insight even more (we had students engage with the notion of physicalisation, where data is encoded through the material and geometric properties of physical artefacts [9]). This is for various reasons. First, creating a physicalisation requires conscious data curation [19] –what is considered important to convey– because the amount of data that can be included in a physicalisation is limited. Second, haptic representations and physicalisations lend themselves less to transporting numerical data (completeness and accuracy), than to conveying an intuitive message, or to a more visceral level of engaging with data. They have been shown to foster self-reflection [17] and appear to trigger deeper connections to data [7, 20]. Lastly, their material nature helps to

Manuscript submitted to ACM

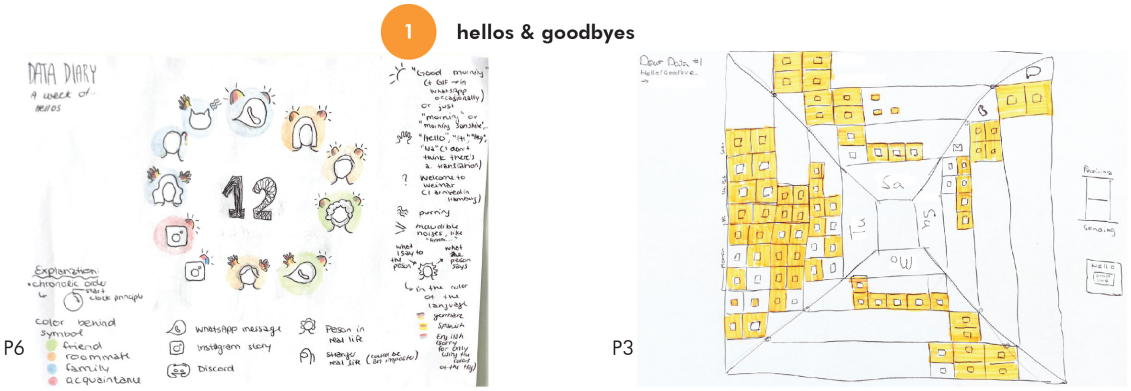

Fig. 1. **Explanation:** Two Data Diaries of Hellos and Goodbyes created for the first assignment in the first iteration of the course. Left: P6 created a rich narrative visualisations representing who, when, where, and via what medium said hello or goodbye (including a purring cat), with colour coding of the type of relationship, the language used, etc. Right: P3 created an analytical visualisation indicating whether an exchange of verbal greetings was complete or incomplete, marking hello's and goodbyes (outer vs. inner box) and P3 as receiver (top box in a row) or as 'sender' (lower box in a row). This shows that some encounters end with goodbyes, without an initial hello and in reverse, sometimes there is no goodbye, or no verbal exchanges at all. (see [18])
Reflection: These two constituted the biggest contrast in the course. Already at this point –when the visualisations were shared in class– students became aware of the different decisions made regarding which data parameters to include, and the fact that decisions on how to categorise an instance often are not clear-cut: Is a 'hmmmh' a verbal or nonverbal greeting? What if a location or person encountered does not fit in the set of categories initially decided on? Do hellos and goodbyes need to match up in style to be considered complete? Moreover, by discussing the Data Diaries, students saw the variety of approaches in representing the same topic, with some visualisations still relying on common visualisation approaches, whereas others created new formats (e.g., P3).

circumvent the imaginaries that data are immaterial, digital, and objective [1, 2, 8] and highlights the connectedness, situatedness, and messiness of data [14, 15].

Our workshop submission is for a large part based on a previous paper discussing our insights from using Data Diaries for a course on data physicalisation, leading to students reconsidering 'what is data' [18]. We augment this here with examples from a second course using a similar set of exercises.

2 DATA DIARY EXERCISES

The Data Diaries are inspired by *"Dear Data"* [12], a year-long project where two authors sent each other weekly postcards with creative data visualisations which illustrate the small, intimate, and interpretive nature of data [1]. The project was documented in a book. We have used the Data Diaries approach three times now in project-oriented teaching, for the first phase where students are to engage with literature as well as to do hands-on exercises that foster creativity and support reflection on the nature of data.

From week to week, project students receive instructions to collect data in their everyday life and create a data representation of these. These exercises move from the more familiar to the unusual representation modality (from the students' perspective): first visual, than physical (using materials), and later haptic or even kinetic. Two courses followed the process described in [18] closely, and another adapted the exercises topic-wise to focus more on personal data. We here describe the process used twice for weekly assignments. Both courses culminated in a larger individual or team creation with a self-chosen data source and topic, which we do not cover here.

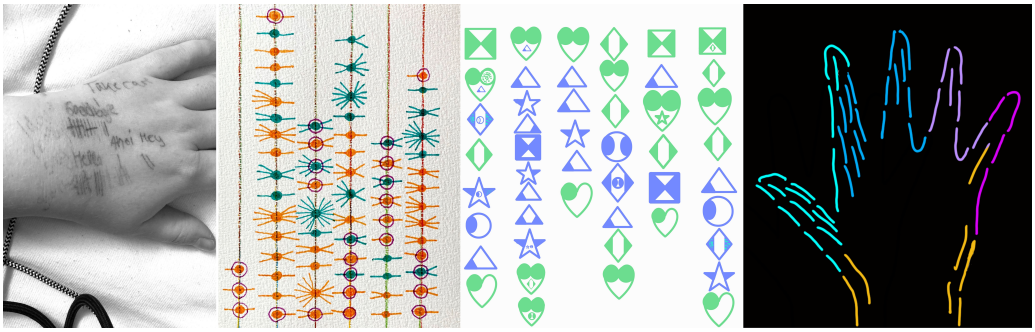

Fig. 2. **Explanation:** Four Data Diaries of Hellos and Goodbyes from the second course. P1 (left) counted different phrasings they used. P2 (middle left) marked for each day (line) the hello (orange) and goodbye interactions (green), adding circles for interactions on the phone. The direction of lines represents whether these came from others or themselves. P3 (middle right) developed an even more sophisticated coding schema. P4 (right) differentiated by colours if hellos and goodbyes were on phone (green and blue), in person (purple and pink) or via texting (yellow - apparently only hellos were exchanged over text), over five days (each finger).
Reflection: In both iterations of the course, we saw different degrees of complexity in encoding, reliance on familiar means of representation from visualisation (colour, shape), and a focus on achieving aesthetic representations. Nevertheless, there is huge variety in what students chose to depict—what is important about hellos and goodbyes? Is the medium of the meeting relevant? Or who you meet? Why does this interest me? Also, the activity of tracking and creating these representations often had students reflect on their social interactions and their behavioural patterns.

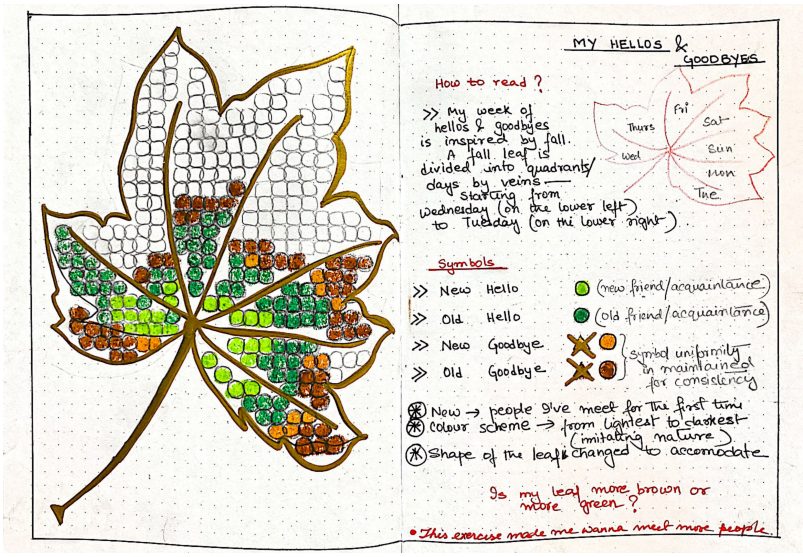

Fig. 3. **Explanation:** Another example from the second iteration uses the poetical metaphor of a leaf in autumn, where each segment is a day. P5 differentiated between hellos from new acquaintances (first encounters, depicted as light green dots) and 'old' ones (dark green), and goodbyes (light vs. dark brown dots, as goodbyes make them sad).
Reflection: This was one of the most poetic representations created, despite relying only on colour. It seemed to foster extensive reflection - P5 asked themselves: *"is my leaf more green or more brown?"* and remarked that the exercise made them want to meet more new people.

- The first exercise has the prompt to document 'Hellos and Goodbyes' (or apologies and compliments) from their life over several days. Similar to *"Dear Data"*, this was not allowed to be a standard depiction (e.g., a bar or

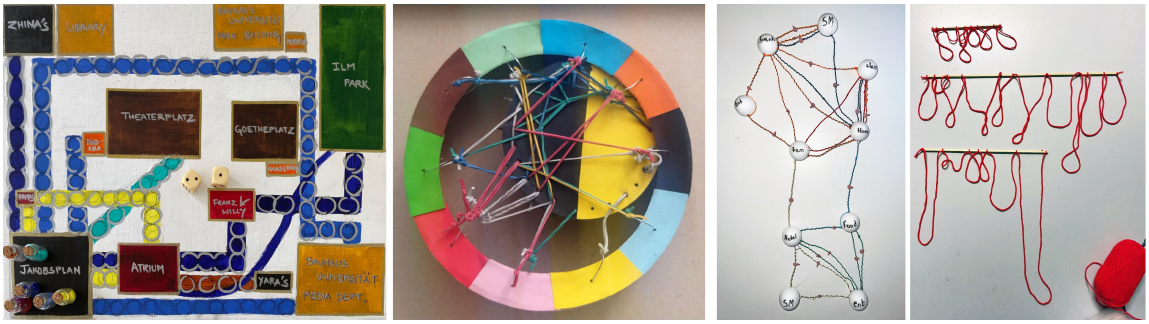

Fig. 4. **Explanation:** For the physicalisation of routes and paths in the second iteration of the course, P5 (left) created a map of their routes through the city over several days (colour-coded by days) and turned this map into a game board (the instructions are omitted). P6 was inspired by scatter plots and musical string instruments, where the object encodes places (the coloured segments around the circle, e.g., friends' homes, pubs, stores, University, train station, work, and home), daytime (the black-yellow disc inside), days of the week (colour of lines). P3 created a physical line diagram (middle left). P6 and P3 both made frequency of particular trips visible by having multiple threads. P2 (right) abstracted their experience by distinguishing time and distance travelled: Each rod represents a day, with the distance travelled encoded in its length. The thread represents the length of time of each individual trip. Time here has become unruly and loopy.
Reflection: Some participants relied on traditional maps, but made them physical, while others invented more abstract representations. By tracking their data, P2 realised that the means of transportation impacted their experience of the route, by influencing the time taken. Therefore, they decided to encode the relationship between time and distance. Writing down all their paths created deeper reflection. For others, the physicality motivated extensions of the representation: for P5, the physical map reminded them of board games, resulting in elaborate game instructions, and P6 began to explore material properties and geometry.

line chart). Instead, students had to figure out new, creative ways to visualise the data. Moreover, these had to be drawn or created by hand.

- Next, data on 'Routes and Paths' are to be collected and represented—in one year this was to be a visualisation, in the next iteration of our course it was a physicalisation (we had learned that we could skip one exercise to move faster towards other modalities). In addition, during the first run of the course, students had to physicalise something in their homes that could be counted and categorised (e.g., number and types of shoes they own).
- The next assignment challenges students to create a haptification—a data representation which communicates data via the sense of touch. In the first iteration of the course this concerned their experience of senses and sensations, in the second iteration they had to represent their energy consumption (either electricity/gas, or mental/personal energy). While many students over the three courses made their biggest learning jump in the haptification exercise, this was also the hardest for all, and some omitted the actual haptics-element (creating something that can be experienced via vision alone).
- For the final exercise, in the first year, students were to represent 'How I spent my time' through movement (resulting in videos or kinetic objects), and in the second iteration, students were to create a sensification [5] of 'lies' (this was open regarding whether to track own lies or other 'sources' of lies. Several students chose topics related to more political and societal issues). Sensifications are physical data representations, which encode data in the behaviour or functionality of the artefact, where data insight is retrieved through the overall experience with the sensification.

Course two had a stronger focus on Data Feminism [2] (through reading sessions and discussions). This, in combination with the mandatory reading of [16] which introduces data activism, resulted in data representations with an

activist message, for example, the paper boats in Figure 7. This sensification represents the 4 countries with the highest CO2 emissions through coloured boats where the paint dissolves in water (thus, polluting it), whereas non-coloured boats represent the countries most affected by the pollution. These boats are punctured, so they sink in the (polluted) water.

3 FROM DATA AS GIVEN AND OBJECTIVE TO DATA AS CREATED AND DATA STORYTELLING

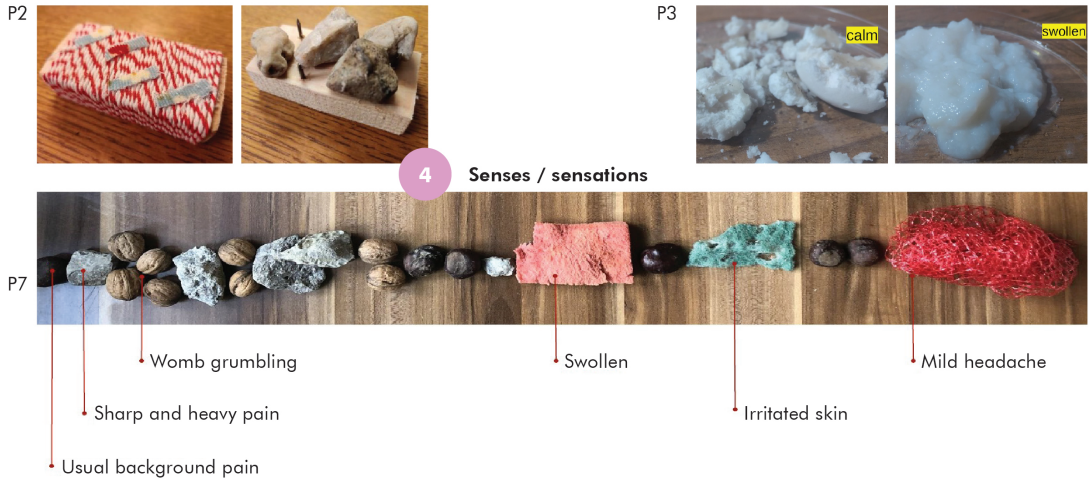

Fig. 5. **Explanation:** Three haptifications from the first iteration of the course, which use the tactile modality to convey data. Top left: Feelboxes created by P2 represent music. Based on the 'feel' of a song (e.g., a rock song feels 'spiky' and rough), P2 created wooden blocks covered with materials. Here we see one block covered in yarn and textile patches, representing a soft song, whereas rocks and nails represent a rock song. Top right: P3's tapioca textures represent bodily feelings, such as 'swollen' as well as emotional feelings, such as 'calm'. P3 transformed the tapioca from one stage to the next, so the previous representation would be gone. Bottom: A timeline of P7's menstrual pain from different materials (e.g., walnuts, stones, a fruit net). (see [18])
Reflection: This assignment marked the first time that participants of this iteration started to change (1) what they tracked and represented as data, and (2) how they represented it. Both the topic of 'Senses and Sensation' as well as the required modality (haptics) moved participants towards exploring subjective aspects of data (e.g., 'how does it make me feel?'), and prevented them from exactly measuring (e.g., how to you quantify the 'feel' of a song?) and representing these data. This made them realise that data are more than numbers and that basically anything could be considered data, when approached as such.

In [18], we have traced how our course participants from the first iteration went through learning stages, where they initially relied on known visual conventions (colour, size) for data representation, were apprehensive of the idea of physical representations, and considered data as an objective and neutral entity, towards embracing the subjectivity of data and learning to create a symbiosis between the story to tell, data, method, and material. On a general level, the Data Diaries helped participants understand the process of collecting and representing data, and primed them to observe the data around them. Key to the these effects was that participants had to collect and represent data themselves.

Creating and discussing Data Diaries highlighted the qualitative aspects of data. As each Data Diary resulted in a variety of approaches in what was counted and how (just hellos and goodbyes, or also the mode of communication, feelings experienced, etc.), the decisions involved in deciding which variables to track, and how to categorise these became visible. This made participants realise that data categories often are not clear-cut. By tracking their own data,

Fig. 6. **Explanation:** P4 (left-most) tracked their 'overall' energy (both physical and mental energy) used to get through the day. They created seven stress-relief pouches from cloth (one per day) filled with rice representing the number of (effort-taking) tasks done. The buttons represent the counting of days. One needs to actively touch the pouches to feel the difference of how much rice they hold. The result is a contradictory product: the busier a day was, the more relaxing squeezing the pouch is. P4 commented that the contradictory outcome of the exercise made the process *"humorous and fun"*. P6 (left-middle) tracked their mental demand on tasks, differentiating pleasure/fun, and stress/concentration, classifying these from 1 to 4. The electric maze represents their level of mental demand per day, enabling somebody else to experience what this day was like. The number of vertices per stretch of metal (one task) shows the mental demand level, where level 1 is a straight wire. P7 (right) tracked how long they ran water, charged devices, took transportation, had the lights or stove on, through which they realised that they consume a lot of electricity and water. Based on this, they created an energy skateboard, mapping water, transportation mode, charging time, and water use to one wheel each. Based on the consumption amount, the respective wheel is slowed down. Trying this out, it turns out that especially water consumption *"slows you down"*.
Reflection: These projects (especially from P7 and P6) are striking in that (albeit one can also just view them) they all invite active interaction with the created object to experience the data in a proprioceptive and kinaesthetic way, not relying on haptics alone. While one can see the electric maze and its wire mess, following the maze has one re-experience the mental demand it represents. In the previous iteration of the course, students had relied mostly on tactile properties of materials, following the literal instruction to *"create a representation that communicates the data via touch"*. We are not sure whether this may be a result of different literature being read or guest lectures that might have inspired students. Still, some students created a physicalisation which can be perceived purely visually, for instance two miniature houses where for one, the light keeps going out, representing energy shortages in the Gaza Strip.

our participants were forced to make decisions on what to count and how they were going to count it. As such, they became aware of the complexity of data and that not everything can easily be quantified.

Interviews with participants from the first course iteration illustrate this change of thinking. One participant explained: *"with the first one [exercise] you still have the old thinking of data how just represents like numbers"* and another explained their initial perception of data as something *"very objective and very efficient"*. Another participant explained their learning: *"There is a human behind every function that encodes the data, so it's always subjective and not only objective like you would expect for the data to be"*. Another explained that they saw a different usage for physicalisations and sensifications: *"some kind of link or little window where you can maybe peep with your eye through and then you can take a look into a life of a different person and this is not about efficiency anymore"* or: *"the goal to give some kind of emotion or give some kind of meaning to the data"*

The tension between quantitative and qualitative increased when creating haptifications or physicalisations. We found that to create physicalisations, our students needed to change their understanding of what data representations (should) do as well as their notion of data. The data could not be solely numeric as the material nature enforces a link to the data's origin. Representations became more poetic and increasingly employed metaphors (including haptic metaphors, e.g., roughness and softness).

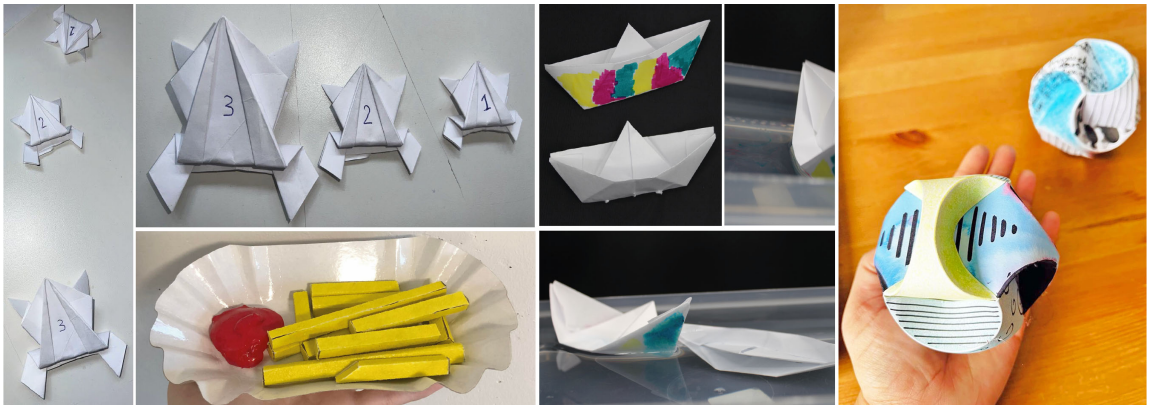

Fig. 7. **Explanation:** P3 (left and top left) created jumping frogs, which represent categories of reasons for lies ((1) to not harm people's feelings; (2) to shorten conversation; (3) self-motivation). The more lies in a category, the higher the frog jumps. P7 (left bottom) created a trompe l'oeil (a lie in itself) of french fries. Each fry represents a lie and is made out of sandpaper. The paper roughness represents the *"degree of the lie"* on a scale from 1-4 (max.). P4 (middle) addressed 'climate hypocrisy' (poor countries are accused of pollution while suffering most from flooding and climate change). They created two paper boats. One represents the four countries with the highest CO2 emissions and is painted with watercolour to depict these emissions. Another represents the countries affected most by this (non-coloured and punctured on the bottom). The boat of the top-polluters will release its colour in the water, depicting polution. The affected countries' boats, on the other hand, will slowly sink in the (polluted) water. P8 (right) represented cyclical harvest poverty, showing the ups and downs of food scarcity from agriculture, where poverty keeps repeating over extended periods.
Reflection: Although this Data Diary concerned lies –which we presumed to be a very personal topic– some participants focused on lies created and upheld by power structures, such as P4's boats and P8's harvest poverty. Furthermore, we received a mix of artefacts which align closer to the definition of data sensification by Hogan [5] (e.g., P3's frogs) and artefacts which can be considered performative objects, where the performance of the artefacts (such as the sinking boats) expresses the story of the data.

Furthermore, they realised that the stories told with physicalisations do not demand accuracy; they might even benefit from a reduction of information [19]. We repeatedly found that students were challenged to curate their data [19], reducing the number of tracked dimensions, and often began to rely more on poetic metaphors.

At this point, they realised that haptifications required them to embrace subjectivity, since it is almost impossible to convey exact numbers through touch. Haptifications and physicalisations were seen to offer the freedom of exploring other values than accuracy, and to enable others to truly re-experience something: *"you can try to map the feeling you're having on the inside, you can try to recreate somehow that experience by touching a surface"* (participant from the second course, which is not reported here).

Through using the Data Diaries in three separate course, we have seen that they are a method which helps to create awareness of data feminist perspectives on data [2]. By tracking your own data, the Data Diaries show that data are messy, need to be seen and counted as data to be interpreted as such, and that there are multiple ways of looking at and tracking the same topic. This highlights that data are capta.

ACKNOWLEDGMENTS

We thank all our participants from the courses for their amazing work. We also thank Hannes Waldschütz who co-taught these two courses.

REFERENCES

[1] Audrey Desjardins and Heidi R. Biggs. 2021. Data Epics: Embarking on Literary Journeys of Home Internet of Things Data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 615, 17 pages. https://doi.org/10.1145/3411764.3445241

[2] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism.* The MIT Press, Cambridge, Massachusetts, USA. https://mitpress.mit.edu/books/data-feminism

[3] Johanna Drucker. 2011. Humanities Approaches to Graphical Display. *Digital Humanities Quarterly* 5, 1 (2011). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html#

[4] Lisa Gitelman (Ed.). 2019. *"raw data" is an oxymoron.* MIT Press, London, England.

[5] Trevor Hogan. 2018. Data Sensification: beyond representation modality, toward encoding data in experience. In *Design as a catalyst for change - DRS International Conference*. Design Research Society, Limerick, 1–15. https://doi.org/10.21606/drs.2018.238

[6] Trevor Hogan, Uta Hinrichs, and Eva Hornecker. 2016. The elicitation interview technique: Capturing people's experiences of data representations. *IEEE Trans. Vis. Comput. Graph.* 22, 12 (Dec. 2016), 2579–2593.

[7] Trevor Hogan and Eva Hornecker. 2012. How Does Representation Modality Affect User-Experience of Data Artifacts?. In *Haptic and Audio Interaction Design*, Charlotte Magnusson, Delphine Szymczak, and Stephen Brewster (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 141–151.

[8] Sun-ha Hong. 2020. *Technologies of Speculation -The Limits of Knowledge in a Data-Driven Society.* NYU Press, New York.

[9] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. Opportunities and Challenges for Data Physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3227–3236. https://doi.org/10.1145/2702123.2702180

[10] Helen Kennedy and Rosemary Lucy Hill. 2018. The feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology* 52, 4 (Aug. 2018), 830–848.

[11] Yanni Alexander Loukissas. 2022. *All data are local.* MIT Press, London, England.

[12] Giorgia Lupi and Stefanie Posavec. 2016. *Dear Data.* Particular Books / Penguin, London. http://giorgialupi.com/dear-data-2

[13] Deborah Lupton. 2018. How do data come to matter? Living and becoming with personal data. *Big Data Soc.* 5, 2 (July 2018), 205395171878631.

[14] Dietmar Offenhuber. 2019. Data by proxy - material traces as autographic visualizations. *IEEE Trans. Vis. Comput. Graph.* 26, 1 (Jan. 2019), 98–108.

[15] Dietmar Offenhuber. 2020. What we talk about when we talk about data physicality. *IEEE Computer Graphics and Applications* 40, 6 (June 2020), 25–37. https://doi.org/10.1109/MCG.2020.3024146

[16] Ricardo Sosa, Victoria Gerrard, Antonio Esparza, Rebeca Torres, and Robbie Napper. 2018. Data objects: Design principles for data physicalisation. In *DS 92: Proceedings of the DESIGN 2018 15th International Design Conference*. Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, The Design Society, Glasgow, 1685–1696.

[17] Alice Thudt, Uta Hinrichs, Samuel Huron, and Sheelagh Carpendale. 2018. Self-Reflection and Personal Physicalization Construction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3173728

[18] Rosa van Koningsbruggen, Hannes Waldschütz, and Eva Hornecker. 2022. What is Data? - Exploring the Meaning of Data in Data Physicalisation Teaching. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Daejeon, Republic of Korea) *(TEI '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 21 pages. https://doi.org/10.1145/3490149.3501319

[19] Hannes Waldschütz and Eva Hornecker. 2020. The Importance of Data Curation for Data Physicalization. In *Companion Publication of the 2020 ACM Designing Interactive Systems Conference (DIS' 20 Companion)*. Association for Computing Machinery, New York, NY, USA, 293–297. https://doi.org/10.1145/3393914.3395892

[20] Yun Wang, Adrien Segal, Roberta Klatzky, Daniel F. Keefe, Petra Isenberg, Jorn Hurtienne, Eva Hornecker, Tim Dwyer, and Stephen Barrass. 2019. An Emotional Response to the Value of Visualization. *IEEE Computer Graphics and Applications* 39, 5 (Sept. 2019), 8–17. https://doi.org/10.1109/mcg.2019.2923483