Open Research Online

The Open University's repository of research publications and other research outputs

Exploring design principles for data literacy activities to support children's inquiries from complex data

Journal Item

How to cite:

Wolff, Annika; Wermelinger, Michel and Petre, Marian (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human-Computer Studies*, 129 pp. 41–54.

For guidance on citations see [FAQs](#).

© [not recorded]

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1016/j.ijhcs.2019.03.006>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Exploring design principles for data literacy activities to support children's inquiries from complex data

Annika Wolff ¹, Michel Wermelinger, Marian Petre

*School of Computing and Communications, The Open University
Milton Keynes MK7 6AA, United Kingdom*

Abstract

Data literacy is gaining importance as a general skill that all citizens should possess in an increasingly data-driven society. As such there is interest in how it can be taught in schools. However, the majority of teaching focuses on small, personally collected data which is easier for students to relate to. This does not give the students the breadth of experience they need for dealing with the larger, complex data that is collected at scale and used to drive the intelligent systems that people engage with during work and leisure time. Neither does it prepare them for future jobs, which increasingly require skills for critically querying and deriving insights from data.

This paper addresses this gap by trialling a method for teaching from complex data, collected through a smart city project. The main contribution is to show that existing data principles from the literature can be adapted to design data literacy activities that help pupils understand complex data collected by others and form interesting questions and hypotheses about it. It also demonstrates how smart city ideas and concepts can be brought to life in the classroom.

The Urban Data School study was carried out over two years in three primary and secondary schools in England, using smart city datasets. Three teachers took part, providing access to different age groups, subject areas, and class types. This resulted in four distinctive field studies, with 67 students aged between 10-14 years, each lasting a few weeks within the two year period. The studies provide evidence that when engaging with data that has not been personally collected, activities designed to give the experience of

¹School of Engineering Science, LUT University, Lappeenranta, Finland

collecting the data can help in critiquing it.

Keywords: data literacy, human-data interaction, smart city, open data

1. Introduction

Society is increasingly driven by data. One example of its use is to inform business decisions, a process that is often referred to as business intelligence. With an increase in data available to businesses, there is a growing gap in the number of employees with the skills to make good use of it. In a policy briefing, Nesta explores this skills gap in detail and proposes ways to address it [1]. Amongst these is a proposal that highlights the importance of initiatives to teach data skills in school and to embed them into other subjects, improving the data literacy of school-leavers and their readiness for the future job market.

Business and employment needs are not the only drivers toward increasing data literacy. Presentation of online content is often decided based on analysis of what users having been clicking through or purchasing online, with the intention to influence the end-users' actions and decision-making. Examples include the recommendations made on shopping sites or entertainment services. Mortier et al. [2] argue that it is important to explore the issue of transparency of how users' data is collected and analysed and how to give increased agency to users who provide data so that they can themselves derive value from it. This is reliant on users having a level of data literacy that enables them to engage with their own data. Beyond this, a white paper of Bhargava et al. [3] highlights the importance of data literacy as an increasingly important skill for civic empowerment. Policy decisions and media reporting are increasingly justified with data, and people therefore need skills to assess critically the accuracy of what is presented to them as fact [4]. One final, yet important, reason for advocating data literacy is that citizens increasingly use data-driven smart technologies to make their lives more efficient, including smart meters, travel apps, or the currently popular 'sharing economy' apps through which people swap knowledge, goods and services. The increasing availability of open data is often mentioned as something that can support 'bottom up' citizen innovation, but this is predicated on citizens having appropriate skills to design around large, complex data sets. However, evidence provided by Janssen et al. [5] shows that this potential is not being reached, and that one of the key barriers is lower levels of data literacy

34 amongst the general population.

35 To understand why this is the case, we turn our attention to what stu-
36 dents are learning in school. Most of the examples mentioned above typically
37 use large and complex data sets and require that people engage with data
38 that they did not personally collect. In contrast, data sets traditionally used
39 for teaching in schools tend to be smaller and are often collected by the stu-
40 dents themselves. Research has shown that when analysing larger and more
41 complex pre-existing data sets students may find it difficult to understand
42 how the data were collected, which in turn makes it harder to interpret [6].
43 In general, skills learned on small data sets may not necessarily scale. This
44 makes an argument for increasing the range of data used to teach data skills
45 in school, which then raises the question how to achieve this in practice.
46 At the same time, the work of Bowler and Acker [7] revealed that students’
47 current understanding of data may be quite limited, for example they might
48 understand the role of data in a scientific inquiry but not necessarily make
49 the connections between their personal data and the different ways it may
50 be used, or abused. Overall, this suggests that students may not be getting
51 the broad data literacy learning that they need at an early age.

52 Despite its importance, there is currently little research that focuses on
53 how to deliver data literacy teaching in the classroom, and in particular
54 teaching that is based on analyzing more complex externally sourced data.
55 This paper addresses this research gap by developing a method that draws
56 on the existing approaches for teaching data literacy for smaller, personally
57 collected data sets, and extends it to larger, externally sourced data. The
58 main contribution is in the synthesis and reframing of existing principles to
59 support the design of data literacy activities so that they can be adapted to
60 this teaching context.

61 This paper reports on an exploratory two-year study in which these design
62 principles were put to the test. Three teachers from three different UK
63 schools took part in this initiative to integrate teaching data literacy skills
64 into both primary and secondary school classrooms. The work described in
65 this paper was conducted in the context of MK:Smart², a large smart city
66 project in Milton Keynes. This project provided an opportunity to develop
67 lesson plans and materials around some less typical data sets that were being
68 collected as part of the project and at the same time to bring smart city

²<http://www.mksmart.org>

69 concepts into the classroom. The lesson plans were used in local schools. The
 70 approach taken was a user-centred ‘research through design’ [8, 9] approach
 71 that fit with the need to be flexible within each school engagement and in
 72 which each classroom engagement generated new knowledge. We discuss how
 73 the findings contribute to the following research questions:

- 74 • What factors influence students’ abilities to ask and answer questions
 75 from the presented data?
- 76 • What is the role of data interaction in facilitating the inquiry process?
- 77 • How does personally collecting a data set changes one’s perspective of
 78 it?

79 2. Background

80 There is no single agreed definition of data literacy and as a consequence,
 81 definitions can vary according to use. Wolff et al. [10] proposed the following
 82 definition to reflect the role of data for innovation:

83 *“Data literacy is the ability to ask and answer real-world questions from*
 84 *large and small data sets through an inquiry process, with consideration of*
 85 *ethical use of data. It is based on core practical and creative skills, with*
 86 *the ability to extend knowledge of specialist data handling skills according to*
 87 *goals. These include the abilities to select, clean, analyse, visualise, critique*
 88 *and interpret data, as well as to communicate stories from data and to use*
 89 *data as part of a design process.” (p. 23)*

90 Deahl [11] proposed that data literacy is: *“The ability to understand,*
 91 *find, collect, interpret, visualize, and support arguments using quantitative*
 92 *and qualitative data.”*

93 Hautea et al. [12] derived what they term *critical data literacies* using a
 94 bottom-up approach that observed young people’s interactions with data and
 95 how this helped them to articulate concerns about privacy and their scepti-
 96 cism around data accuracy, for example when they spotted inconsistencies
 97 in the data presented.

98 Despite this diversity of focus, there is a growing convergence on the idea
 99 that data literacy is more than simply learning a set of technical skills, such
 100 as how to read bar graphs [13, 14], work with maps [15] or use data for
 101 prediction [16]. While these are essential skills and worthy of study, other

102 initiatives have taken a broader view of what is data literacy and how to
103 develop it, especially within formal school education.

104 Among these approaches, several have focused on supporting data-driven
105 inquiry. These include the work of Lee, Drak and Thayne [17] who used
106 quantified self data to engage students with familiar personal data and then
107 prompted them to drive their own inquiries from the data. The Local Ground
108 project [18, 19] developed a geo-spatial data collection tool that students
109 could use in geo-spatial data-driven inquiries. Dasgupta and Hill [20] sup-
110 ported children to drive their own inquiries from data and to create their own
111 visualisations, using the Scratch programming environment, which many chil-
112 dren are already using in school for programming. However, certain aspects
113 of the inquiry process are found to be problematic, in particular how to link
114 questions and data [21, 22].

115 Complementary to this, other approaches put the focus on the ability to
116 use data for civic empowerment. These include the City Digits project [23]
117 that aimed to teach data literacy skills to school children by encouraging
118 them to investigate social issues in a local, urban context. Also, the Data
119 Murals project [24] brought together a community to build an artwork that
120 reflected their data explorations with data from and about their neighbour-
121 hood. Anslow, Brosz and Maurer [25] explore the potential of datathons for
122 building data literacy, which bring together students and members of the
123 community to solve problems.

124 Also gaining traction is a STEAM based approach. For example, D’Ignazio
125 [26] focuses on approaches that support non-experts to learn important skills
126 for framing problems around complex data through creative, rather than
127 technical, activities.

128 Underpinning these, a number of principles to support data literacy learn-
129 ing have been proposed. These include the principles of data informed learn-
130 ing by Maybee and Zilinski [27] which propose that:

- 131 1. New ways of using data must build on students’ prior experience.
- 132 2. Learning to use data should occur at the same time as learning about
133 a disciplinary subject.
- 134 3. Learning should result in students becoming aware of new ways of using
135 data as well as developing new understandings of the subject being
136 studied.

137 Srikant and Aggarwal [16] proposed and tested these principles:

- 138 1. Use a full data cycle.

- 139 2. Make the data set relatable (e.g. about themselves).
- 140 3. Avoid pre-built data sets, but get students to do the task of data col-
- 141 lection and entry themselves.
- 142 4. Reduce problem complexity (for example, if teaching predictive models,
- 143 use only 2 categories).

144 Taking a slightly different approach, Bhargava and D’Ignazio [28] propose
145 a set of design principles to use while developing tools to support data literacy
146 learners, suggesting tools should be:

- 147 1. Focused, to do one thing well.
- 148 2. Guided, to help get the learner started.
- 149 3. Inviting, to appeal to the learner, maybe using data on a relevant or
- 150 meaningful topic to the learner.
- 151 4. Expandable, offering paths to deeper learning.

152 The data literacy initiatives described have one thing in common, in that
153 they focus on the use of data that is collected by the students themselves. As
154 discussed, while clearly an essential skill, this does not necessarily translate
155 to skills for dealing with externally sourced data [6]. Similarly, none of the
156 data literacy design principles address this need, in fact the principles of
157 [16] actively steer away from this, suggesting the students only engage with
158 personal data. We instead propose to harness these same principles to *help*
159 students engage with large, external data sets, through a small adaptation to
160 a principle related to *personal data collection*. These principles are described
161 in the following section. At the same time, there is little discussion in the
162 literature of how such principles can be applied in practice, or how tools have
163 been designed using principles for tool development described by [28]. We
164 therefore show how these principles have been used to guide the co-creation
165 of a set of lesson plans and the design of new tools that *complement* them,
166 and then we explore how they are used in real classroom settings.

167 3. Data Literacy Activity Design Principles

168 We propose the following set of principles to support the design of activi-
169 ties for teaching data literacy, which synthesises the existing principles found
170 in the literature. The main contribution is in the adaptation of a personal
171 data collection principle (P6) to show how personal data collection can be
172 used to complement interpretation of existing data, rather than to be used
173 *instead* of it:

174 **P1 Inquiry Principle:** Follow an inquiry process to scaffold the data
175 analysis. Lead the students first in a guided inquiry, from which follows an
176 open inquiry when students are more familiar with the data and the approach.
177 [16, 28].

178 **P2 Expansion Principle:** Start from a representative snapshot of a
179 small part of the data set and expand out, rather than starting with the full,
180 large data set and focusing in. This aims to help students' more easily relate
181 questions to data [22] and to be expandable and offer paths to deeper learning
182 [28]. It aims to provide students the opportunity to orient themselves within
183 the data, before navigating across it, e.g., through time and/or space and/or
184 some other dimension of the data.

185 **P3 Context Principle:** Teach in a context the student understands,
186 using data that is from their own environment, either local to them, or else
187 relating to them in some other way [27, 16, 28].

188 **P4 Foundational competences principle:** Focus on developing foun-
189 dational competencies rather than practical skills, for example how to ask
190 'good' scientific questions from data [21, 22].

191 **P5 STEAM principle:** Take a STEAM approach by working collabo-
192 ratively on creative activities alongside practical ones [26, 24].

193 **P6 Personal Data Collection Principle:** Students should engage
194 with data they have collected themselves. When students are analysing an
195 external data set, they should be given additional activities that support
196 them in understanding what it is like to collect that type of data. This is
197 to support them in contextualising and interpreting the data external data,
198 which according to [6] they may otherwise struggle with.

199 The remainder of the paper describes how these principles have been
200 used in practice to guide creation of lesson plans based around data collected
201 within a smart city project. We focus particularly on evaluating the use of
202 principle P6.

203 4. Iterative Design of Lesson Plans

204 The overall methodology can be categorised as *research through design*.
205 This is a method in which design practice is applied to the creation of arte-
206 facts as a way of exploring solutions to problems, especially 'wicked problems'
207 [8, 9]. In research through design, new knowledge is constructed by undertak-
208 ing activities associated with design, such as iteratively creating and testing
209 prototypes to understand and solve a problem and to act as a focal point for

discussion by making interactions observable. This approach is fairly similar to that taken by data literacy initiatives, such as City Digits [23] and Data Murals, [24], though they are not necessarily framed that way. In our case, the research through design process was focused around the interpretation and use of the activity design principles to create lesson plans to teach data literacy skills and support interaction with smart city data and what we could learn by putting these into practice and through the iterative improvements to lesson plans over time. The relation between the design decisions and the design principles are highlighted throughout the text describing the lesson plans.

We adopted a user-centred iterative design approach with a small group of teachers. There were a number of stages: scoping; identifying potential data sets; drafting lesson outlines; creating an initial set of activities and lesson plans; introducing technologies. Each stage is described in turn.

Scoping: This first stage, which aimed to set boundaries on the types of activities that could be proposed, occurred prior to any engagement with schools. In this stage the decision was made to a) build activities that could be deployed using standard classroom equipment, technologies or software (e.g., iPads, desktop computers, web browsers) and b) build lesson plans from existing data sets, rather than being dependent on capture of data by students, e.g., through sensor technologies. This was in order to keep the initial focus on how to design learning experiences with these external data sets.

Identifying data sets: The second stage involved identifying a number of data sets that were available and could potentially be used for teaching. This resulted in a pack showing representative ‘snapshot’ visualisations of a small part of a number of data sets with some generalised lesson outlines that were broadly speaking agnostic of any particular teaching approach (e.g., inquiry-based, collaborative learning). These lesson outlines identified the types of questions that could be answered by the data, but did not propose any activities or constitute a lesson plan. They were intended to help teachers to understand the data, as it would be unfamiliar to them, and to act as a starting point for discussions. The chosen data sets were all related to the topic of *renewable energy*. They included smart meter data and data on solar energy potential for a number of houses in the city. They were at the time being used within smart city research into load shifting (trying to change typical patterns of energy use to times when overall demand for energy is lower) and in identifying new opportunities for solar installations

248 or community energy solutions.

249 **Lesson outlines:** The third stage involved teachers from two schools,
250 one primary mathematics teacher and one secondary science teacher, who
251 had expressed an interest in using data from the smart city project in their
252 classrooms. Each was invited to discuss the data sets and lesson outlines and
253 how they could be formed into lesson plans. The possible use of an inquiry-
254 based approach for teaching was also discussed. The teachers confirmed
255 that these were not typical data sets used in teaching and were keen that
256 students would get some experience in handling these different types of data.
257 While the teachers came from different subject areas, the topic ‘data inquiry’
258 was seen to fit quite well in either mathematics or science, and ultimately
259 the subject area did not play a big part in shaping the lesson plans. The
260 secondary school science teacher was very familiar with an inquiry approach,
261 as used in science, and was keen that this would be the approach used with
262 the data.

263 **Teaching activities and lesson plans:** Through these discussions,
264 the initial set of teaching activities and lesson plans was created, based on
265 the principles P1-P6 described earlier. Tasks were adapted for each specific
266 school context, based on the recommendations of the class teacher, so that
267 the experience would align with what the students had been learning and be
268 suited to their overall abilities. This allowed us to gain a better understanding
269 of what the overall differences might be between schools and age groups, but
270 ruled out a controlled approach to evaluation, across different school settings.
271 These lesson plans are described in the next section.

272 **Introducing technologies:** The first trials were conducted using pa-
273 per materials. Later trials introduced technologies to support interaction
274 with the data, being focused on only simple functionality [28] to support
275 key aspects of the task (as identified through first trials) and following the
276 *expansion principle* (P2).

277 5. Lesson Plans

278 For each lesson plan, we describe: a) the overall aims of the lesson and the
279 data set on which it was based, whether it was an existing data set or collected
280 by the students for the purpose of contextualising one of the data sets; b) the
281 activities undertaken with the data and how they were related to the design
282 principles; c) the intended outcomes. The activities were used in various
283 configurations across four separate field trials. The configuration was decided

284 based on several meetings with the teacher. It should be noted that while
 285 there was never any need to adapt materials based on the classroom subject,
 286 the introduction that was given to the class prior to starting activities was
 287 different in each case, based on students' prior knowledge. These general
 288 introductions are not discussed further in this paper. Some other lesson plan
 289 variations were necessary due to the age of the students and also based on
 290 developments that happened in technology during the period of the project.
 291 These variations and their reason are indicated.

292 5.1. Lesson Plan 1 (LP1): Smart Meter Energy Data

293 The aim of this lesson was to show, through data, how energy consump-
 294 tion and generation from solar panels did not always match if people were
 295 not typically at home during the day when solar energy was being produced.
 296 This lesson used smart meter data from approximately 70 houses. For each
 297 property, students had access to data about: a) whole house consumption;
 298 b) individual appliance consumption; c) generation of solar energy. The ex-
 299 ample (figure 1) shows whole house consumption for one day in March. This
 300 data was anonymised, but it came from the same city that the students in-
 301 habited and this was conveyed to students to help them to contextualise the
 302 data (P3).

Figure 1: Smart meter data showing whole house consumption in one day

303 *5.1.1. LP1 Activities*

304 Students followed an inquiry process, based on posing questions from the
305 data set (P1). The guided inquiry stage started with a snapshot of data (P2),
306 as in figure 1, and some questions to answer from it. These asked when was
307 most or least energy used and also prompted students to tell a story about
308 the people living in that property, based on how they were using energy.
309 Students worked in groups on all activities.

310 After familiarisation with the data, the next stage prompted students
311 to explore the wider data set (P2), for example, answering questions about
312 whether all houses showed the same pattern, or if the patterns varied at
313 different times of year. There were variations in how this stage was delivered,
314 which were tailored based on the age of the students and the development of
315 technologies over the course of the project. The variations were as follows.

316 **Guided:** Students were guided using existing questions. This was used with
317 younger students.

318 **Guided, then Open:** After the guided inquiry, students asked and an-
319 swered their own questions. This had two stages, a brainstorming stage
320 where students posed question and discussed them as a class, then a
321 refinement stage, where they chose just one or two questions to follow
322 up from the data (P4). This was used with older students.

323 **No technology:** Students worked from paper. Data was curated, either
324 into further snapshots (guided activities) or based on the refinement
325 stage, raw data was curated for students to explore one week later
326 (open activities).

327 **With technology:** Students could ask and answer questions rapidly through
328 the data browser (open activities). The data browser supported the se-
329 lection of different houses. It followed approximately the design shown
330 in the Balsamiq mockup in figure 2, with the exception that to config-
331 ure the interface to view different houses required to first submit the
332 house numbers and then select the rest of the attributes (time period,
333 data).

334 *5.1.2. LP1 Outcomes*

335 The intended outcomes were that students would be able to use the data
336 to identify common patterns in energy consumption and to see how these
337 differ by day (e.g., weekday/weekend), household or time of year.

Figure 2: The mockup from which the Interactive Smart Meter Data Browser was created

5.2. Lesson Plan 2 (LP2): Potential for solar energy production

The aim of this lesson was to demonstrate, through data, that houses differ in their potential for producing solar energy, based on the direction they face and the size and pitch of their roof. This lesson used data that was derived from aerial photography, using LiDAR technology. This data set showed the potential energy production by installing solar panels on each building within the city. The data came from the local area and students were able to look at their school and their own houses (P3).

5.2.1. LP2 Activities

Students followed an inquiry process (P1) where they answered questions from the data. As in the smart meter example, the guided inquiry stage started with a representative snapshot of data (P2) from which they could see roughly the size of roofs and where a solar panel might go, colour coded according to whether it was predicted to give a low or high solar yield (figure

352 3). Students worked in groups on all activities. Students were prompted to
 353 answer the following questions:

- 354 • Which house is best for fitting solar panels to? Which is the worst?
- 355 • Look at the houses on the map, why do you think these are good/bad?

Figure 3: Solar potential data set

356 There were variations in how this stage was delivered. For LP2 there was
 357 no planned open inquiry stage as this was delivered only to younger students.
 358 Instead, the variations of the guided inquiry were:

359 **No technology:** Students were given a printout of the map and the snap-
 360 shot area was an estate close to their school that they were all familiar
 361 with. The associated data could be found from a table from which they
 362 could look up each property by the ID and find data about the solar
 363 potential, orientation, size and pitch of roof as well as the estimated
 364 cost of the panel.

365 **With technology:** students used an interactive map that allowed them to
 366 zoom, pan, search by postcode, select the satellite or streetmap layer,
 367 and click on an area of the map to view data. This is shown in figure 4.

368 Through this, they could navigate across the city and ask and answer
 369 their own questions from the data, thus following the expansion princi-
 370 ple (P2). In the guided inquiry stage, these students entered their own
 371 postcode to select a region of houses from their own area from which
 372 to answer the above questions.

Figure 4: Urban Data School Solar Potential lesson plan showing the Interactive Solar Data Set

373 5.2.2. LP2 Outcomes

374 The intended outcomes were that students would: a) understand how
 375 roof size, pitch and direction affect solar yield; b) understand the difference
 376 between interpreting data from the map and from a table (e.g., ability to
 377 see things blocking solar panels compared to ability to do statistics); c) find
 378 errors in the data and understand that data can be flawed.

379 5.3. Lesson Plan 3 (LP3): Be a LiDAR device

380 The aim of this lesson was to provide students with the experience of
 381 capturing data by aerial survey. This activity is based on the personal data
 382 collection principle (P6).

383 5.3.1. LP3 Activities

384 Students were shown the principles of using light to measure distance,
 385 with the help of a portable laser measuring tool. Students then worked in
 386 groups and started by building their own house from plasticine onto which
 387 they marked a grid of 1cm by 1cm (figure 5). This follows the STEAM
 388 principle (P5). They then used home made rulers to measure the height
 389 of each square, transferring their data onto a sheet of paper. Groups then
 390 swapped their sheets, to see if they could understand the shape of the house
 391 from the data alone.

Figure 5: Steps for creating the plasticine house with grid

392 5.3.2. LP3 Outcomes

393 The intended outcomes were that students would understand how LiDAR
 394 data builds a picture of a landscape. They should also understand about data
 395 resolution and how this affects accuracy and the trade off between processing
 396 large data sets and having accurate measurements. A further aim was to
 397 improve their general understanding of how the data for the solar yield of
 398 roofs was created.

399 5.4. *Other activities*

400 We have described three lesson plans that were constructed and used
401 across the field trials. We omit some activities that do not contribute to the
402 later discussion and where, on the whole, the findings are reported elsewhere
403 [29, 30]. One activity that should be mentioned is 'be your own smart me-
404 ter', which encouraged students to collect their own energy data according
405 to principle P6 and then to create novel visualisations from it. This was
406 conducted each time in conjunction with LP1 to contextualise the smart me-
407 ter data. The decision to exclude it was to reduce the amount of results to
408 report - instead we have opted to discuss this principle in terms of LP2 and
409 its complement LP3.

410 6. Methodology

411 We recruited three teachers to participate in four ethnographic field stud-
412 ies using the developed lesson materials within their classes. One teacher par-
413 ticipated in two separate field studies in two different years of the project.
414 Each field study comprised two or three classroom sessions in which stu-
415 dents undertook the activities, usually at one week intervals. There was a
416 constraint in recruiting schools, in that they needed to be in the geographic
417 location covered by the data sets. Teachers were recruited through personal
418 contact.

419 The constraints and method of recruitment meant that we ended up en-
420 gaging with teachers of differing ages, subjects and abilities. Each field study
421 was therefore adapted to align with the requirements of the teacher and their
422 class. This process was led by the teachers, who were invited to select only
423 activities that suited them and to adjust the design of these selected activi-
424 ties then decide how the teaching sessions would be delivered and who would
425 lead: either the teacher, the researcher or a co-led session between teacher
426 and researcher. In the classroom, all activities were undertaken by students
427 in groups of 2 or more.

428 Evaluation at the end of each field study led to incremental improve-
429 ments to the design and delivery of lesson plans, also taking into account
430 the adaptations required by the teacher for the following field study. In ad-
431 dition, the technologies to support teaching were developed and used in the
432 final two studies. This need for flexibility lent itself to a long-term qualita-
433 tive approach to evaluation, rather than controlled studies where it would be
434 possible to collect quantitative data.

6.1. Data collection and analysis

Data was collected for the purpose of refining the approach in a future iteration and also with a focus on assessing the students' ability to link questions to data and to start to form their own inquiries. Data was collected from students in both primary and secondary schools. The total age range of students participating in activities was between 10 and 14.

Each field study was observed by one or more Participant Observers (POs), who recorded videos or took photographs and made notes both during and after the sessions. Participant observation is useful for understanding how people relate, to each other and to task materials, and to identify future questions to be answered [31]. The observation procedures were discussed between observers beforehand. POs were tasked with noting when students needed help, in identifying parts of the lesson plans that caused problems and most importantly any evidence that students were thinking beyond the initial activities and posing their own questions from the data. POs were also tasked in noting down the number of students engaged in tasks and how they formed into groups. The level of participation of the observers varied from co-leading the session to supporting students in practical group work activities. As POs were busy during the sessions, the main data was captured in a summary that was written up as notes immediately after each session. Where practical, verbatim quotes of students were captured at the time, but this was not systematic.

At the end of each field study, the photographs, verbatim quotes and PO summaries were combined to create a single narrative about what was happening in the session, focusing on what problems were encountered and what questions did students ask.

In two field studies that were conducted with older, secondary school students, additional data was collected directly via worksheets and from classroom materials (such as post-it notes). This captured the questions that students asked from data at different points throughout the activities. A qualitative coding of this data to assess the questions for *answerability from the data* was undertaken by the first author, who had expertise with both the data and its use in research. It was verified by a second researcher, leading to some adjustments until a consensus was reached. This process aligns with the process undertaken by [21]. Both an inductive and deductive approach was taken to the coding. In this process, some initial categories were suggested and used to guide the first coding, then these were refined based on the analysis of each question.

Due to the longitudinal nature and slightly differing focus in each field study, the data collected was different in each case which made controlled experimentation difficult. However, each individual classroom session yielded rich data from observations and working materials.

7. Results

This section is structured according to the research questions listed in section 1. For clarity, results that do not contribute to this discussion will be reported on only minimally, or left out altogether.

There were four field studies; a total of 67 students took part. These are shown in Table 1 in the order in which they were conducted, approximately 6 months apart each time.

Id	Sessions	Year(age)	Pupils	Subject	Lead	POs	Activities
FS1	2	5 (10-11)	12	Maths	co-led	1	LP2 no tech
FS2	3	9 (13-14)	17	Triple science	teacher	2	LP1 no tech
FS3	2	7 (11-12)	25	Geography	researcher	1	LP1 with tech
FS4	2	5 (10-11)	13	Maths	co-led	2	LP2 with tech then L

Table 1: Field Study details

Figure 6 summarises the findings from across the four field studies and details how they are used to answer the research questions. These findings are expanded upon in the remainder of the results section.

7.1. Answering RQ1: What factors influence students abilities to ask and answer questions from the presented data?

Lesson plan 1 was designed to follow standard inquiry processes (P1), starting with a guided inquiry and then moving to a more open inquiry with older students. Following the foundational competence principle (P4) and knowing that students may struggle in particular to relate questions and data - which is an important part of the inquiry process, especially an open inquiry - the following results explore the extent to which this was supported through the activities. The focus is on on a comparison between the FS2 and FS3 brainstorming activities of Lesson Plan 1 (see figure 7). This is the start of the open inquiry stage and it took place after all students had completed the guided inquiry from the snapshot of the data (first part of LP1). This

Research Question	Methods	Related Field studies	Participants: No. (age)	Activities	Analysis	Main findings
RQ1: What factors influence students' abilities to ask and answer questions from the presented data?	Comparison between 2 field studies with qualitative data collection through worksheets, classroom materials and participant observation	FS2 FS3	17 (13-14) 25 (11-12)	LP1 open inquiry stage, with smart meter data (both studies)	Categorization of questions framed by students in open inquiry	Younger students found it harder to ask questions directly from the data (see section 7.1)
RQ2: What is the role of data interaction in facilitating the inquiry process?	Comparison between 4 field studies with qualitative data collection through observation	FS1 FS2 FS3 FS4	12 (10-11) 17 (13-14) 25 (11-12) 13 (10-11)	LP2 (no tech) LP1 (no tech) LP1 (with tech) LP2 (with tech) Technology was in the form of an interactive data browser for a) smart meter data b) solar panel data	Narrative construction based on participant observation of students doing the activities	Students who were able to interact with the data were observed to start following their own inquiries, even when not prompted to by the worksheets (see section 7.2)
RQ3: How does personally collecting data changes one's perspective of it?	Single field study with qualitative data collection through observation	FS4	13 (10-11)	LP2 solar panel task, followed by LP3 LiDAR data collection task	Narrative construction based on observation and video recordings	Students become more critical of data when they gain experience in collecting it (see section 7.3)

Figure 6: Summary of results

relates to the categorisation of questions that students made in this stage (see row RQ1 of figure 6).

The question categories that were obtained through coding were as follows. We include also their alignment to the question categories used by Shelley et al. [21]. We have included the 'not answerable' category here, as this was originally suggested prior to coding taking place. However, this category was not needed in the end.

C1 Smart meter questions (completely answerable): students pose a question that can either be answered directly from a further analysis of the smart meter data, or where the further analysis could give enough information for them to form a reasonable hypothesis (that may then lead to further information being needed to verify).

Figure 7: Some students placing their brainstorming questions onto a whiteboard

511 **C2 Supplementary questions** (conditionally answerable): students
512 would require further data or information to answer the question, but this
513 answer would help to interpret findings from the smart meter data.

514 **C3 Topic questions:** questions students have that aid general under-
515 standing of the topic, but are not directly related to the smart meter data.

516 **C4 Validity questions:** students query the validity of the data.

517 **C5 Not answerable:** the question is out of scope for both the topic and
518 the data.

519 It should be noted that, in categorising questions, the goal was to assess
520 the ability of the students to frame questions around the smart meter data
521 for which they could offer a line of reasoning by which their proposed analysis
522 may provide an answer to their question, rather than to judge the quality
523 of this reasoning. Hence, the first category combined questions that could
524 be answered from the smart meter data and those for which the analysis
525 could lead them to form a hypothesis that might then need verification from
526 additional data. Therefore, in completing the categorisation, attention was
527 paid to the explanations given by the students either in their workbooks or
528 in discussion with the teacher or researcher (which were recorded as obser-
529 vations). Where students could offer a plausible explanation of what they
530 would be looking for from the data and how this would relate to the ques-
531 tion, the question was placed into the first category. To give an example,
532 one teacher queried how students would tell from data if there were a young
533 family in the house. A student offered an explanation that the “mini-spike in
534 the energy data could indicate a young family having to heat food, put music
535 on”. With regard to the possibility of visitors being in the house between
536 8:00 and 12:00, a student suggested they “could check whether this happens
537 every day by looking for a spike on other days”.

538 Next, we counted the questions that appeared in each category. We
539 did this separately for FS2 and FS3, to enable comparisons between them.
540 Figure 9 lists all of the questions in the FS2 session and how they were
541 categorised. Additionally, we know whether these questions were selected for
542 further analysis in the refinement stage and by how many students. This
543 information is also presented in the table (it will be discussed in more detail
544 in the next section). It should be noted that some students did not specify in
545 their workbooks which questions they had selected for the further analysis,
546 whereas some students decided to write down new questions that had not
547 been presented by the whole class in the brainstorming stage.

548 FS2 students posed a total of 18 questions across the two stages (brain-

	Brainstorm (shaded questions added at refinement stage)	selected
C1	Does the house have the same pattern every day? we would need another 6 more graphs to compare	2
	Do you have a young family?	3
	Were there visitors in the house between 8am-12pm?	1
	Is it a house full of adults or a family?	2
	What is the average amount of energy consumption of that day shown in the graph? (per hour) (House 4)	0
	If there's no children, do the adults who live there work on a schedule (e.g. 9-5) or work irregular shifts?	3
	Did the whole street's power go, not just that single house?	7
	Is there the same amount of energy used on weekends and weekdays? Do different households use the same amount of energy? Is there always power cuts every once in a while? If answer yes or no – please explain why?	2
	Does this house have the same energy consumption to other houses?	1
	Are the adults employed or unemployed?	1
	What season is it?	1
	How much energy is used at this busy time of year?	1
	When does the electricity usage nationally peak? I chose this question because you can see when a household is most active.	1
	Who might be living in House 4? And how many?	2
C2	If family home, is child(ren) home-schooled?	0
	Where is the location (e.g. countryside or city)?	0
	What is the day? A weekday or weekend? (*note, the smart meter data only indicates dates, not weekdays)	1
C3	What is your smart meter data?	0
C4	- None asked	

Figure 8: Questions asked in FS2 related to energy consumption

storm and refinement). The majority of questions in the refinement stage were chosen from those where the answer was in the data (25) compared to from additional data (3) or general topic (none) indicating that their understanding of how to select good questions was improving through the class discussions and use of technology to interact with the data.

Figure 10 shows the questions asked by students in FS3. They did not formally write down questions for the refinement stage, so this information is missing from the table, but is discussed (based on the observations) in the next section 7.3. FS3 students asked a similar number of questions as FS2, despite a greater number of students (25 students compared to 17). In both field studies, the students worked in groups of two or three.

As described in row RQ1 of figure 6, the notable result is that FS3 students asked fewer questions of the data and more about the data, indicating some difficulty in framing these types of questions. For example, FS3 stu-

	Brainstorm
C1	How often do power cuts happen?
C2	How much energy does the average family use?
	What household item uses the most energy?
	How is it possible to not use electricity in a day on the weekends?
	Why do the family have no electricity in the middle of the night?
	Why do they use less during the middle of the day?
C3	Would energy ever run out?
	How much money averagely spent on energy?
	Why do you have to use a smart meter?
	What happens when too much energy is formed, does the smart meter warn them?
	How much is the smart meter?
	How do smart meters measure microwaves' or toasters' energy use?
	Where else other than homes do you get smart meters?
C4	Does the smart meter always collect energy every day, hour and second. Does it ever stop working?

Figure 9: Questions asked in FS3 related to energy consumption

dents noticed that less energy was being used in the middle of the day and asked why. On the other hand, FS2 students framed much more specific questions that could be answered by looking at more data from the smart meter data set, such as “Does the house have the same pattern every day? We would need another six more graphs to compare.” FS3 students also had many more questions that would aid their general understanding of the topic (C3). The differences between FS2 and FS3 were the age of students (FS3 students were approximately 2 years younger) and the lesson’s subject (science in FS2, geography in FS3).

Overall, the students were able to:

- frame new questions of the wider data set after initially focusing on just a very small part of it;
- create plausible explanations of their findings - even if sometimes the explanations were not the only possible ones and even though they were often not verifiable without additional information.

7.2. Answering RQ2: what is the role of data interaction in facilitating the inquiry process?

This section compares the lesson plans, LP1 and LP2, undertaken firstly without technology and secondly with the use of an interactive data browser

582 - in each case, by a different set of students at a different point in time.

583 7.2.1. *Technology use in FS2 and FS3*

584 FS2, in which LP1 was conducted without the use of technology, is de-
585 scribed in the previous section 7.1. This section focuses on the refinement
586 stage in FS3, in which students were able to ask and answer questions rapidly
587 using an interactive tool in which they could select the smart meter energy
588 consumption data for a time period and a house in which they were in-
589 terested (see figure 6, RQ2: comparing LP1 with and without technology).
590 They could also view data at the appliance level. This data came from smart
591 plugs, which could be configured by each individual household.

592 The data in this stage is based on the observations of the participant
593 observers (POs), as these students did not have time to write their findings
594 in the book. Observations were based on what students were looking at and
595 on summaries of the conversations that students in a group had with each
596 other, or with the PO. Any interpretations presented in these results are
597 based on the interpretations made and written by the POs at the time.

598 The observers noted that students could quickly grasp the meaning of
599 the graph without any help at all, and were starting to answer questions
600 immediately about the times of highest/lowest energy use, as well as start-
601 ing to propose theories for what caused them (see the findings for RQ2 in
602 figure 6). Students could also easily identify the relationship between the
603 graph and daily life activities of the occupants of the houses. This was evi-
604 denced through the stories that students told about what they thought was
605 happening in the house, based on the data. In this case, students tended to
606 focus on questions that compared either a single property or appliance across
607 different time periods. One explanation for this is that the interactive tool
608 made selection of appliances and time periods easier than changing to view
609 a different property. Although it is not clear from the mock-up in figure 2,
610 there was one additional button to press to select the data set of a different
611 house.

612 The queries and explanations were analysed and categorised using the
613 same process as for the questions (section 7.1). These questions (by nature of
614 the task) all belonged in category *C1*, in that they were *completely answerable*
615 from the data so the aim was to undertake a deeper analysis of the types
616 of questions that fell within this category to show what students were most
617 interested in. This analysis revealed that questions fell broadly into two
618 categories. These are now discussed, with some representative examples of

619 explanations.

620 **Comparing a single property at different times:** One group found a
621 reduction in energy consumption at Christmas hypothesising that the family
622 may have spent Christmas elsewhere. Another group focused on anomalies,
623 first discussing possible reasons for a zero value, including the possibility
624 of a power cut. Another student in the group said a power cut would last
625 longer, so perhaps a fuse had gone in the house and the person had woken
626 up and gone and flicked the fuse box back very quickly. Another student
627 thought that perhaps it was a key meter. This same group also noticed two
628 spikes in the data, which they discussed with the researcher, leading to the
629 explanation that perhaps the smart meter was in error.

630 **Comparing a single appliance at different times:** One group was
631 looking at TV consumption and found that the family had suddenly stopped
632 using the TV. They speculated that the TV was broken, but could not think
633 of any other reason, for example, they did not know that the smart plug
634 might have been moved and used to monitor something else. When told
635 this, they decided that this was a more likely explanation.

636 7.2.2. *Technology use in FS1 and FS4*

637 In FS1 students undertook activities related to LP2 (solar potential) using
638 paper-based maps and associated data sets given in a printed table (see RQ2
639 figure 6). The aim was that students would understand how direction, roof
640 area and pitch contributed to solar yield. Students worked in groups. At the
641 end they presented their findings. Their conclusions after engaging with the
642 task were:

- 643 • “If the house [roofs] are slanted then they have the most chance of
644 getting the most electricity.”
- 645 • A 4-sided roof would be “harder to put solar panels on, because some-
646 times the sun doesn’t come from that side.”
- 647 • “It’s best if the solar panels are facing south, because that’s the direc-
648 tion of the sun in the day.”
- 649 • “Even if you buy these really big expensive solar panels, it might not
650 make much of a difference - it might be a waste of money.”

651 Overall, these answers reveal that students had picked up important prin-
652 ciples about solar panels through interpreting the dataset. These include that

Figure 10: Interacting with the solar map

the roofs must face a certain direction and be slanted to get the most sun. They had also begun to understand some of the cost implications.

In FS4, students followed the same set of activities, but they used an interactive version of a map showing solar potential of all roofs in the city (see RQ2 figure 6). Students, working in groups of two or three, first undertook the guided inquiry stage based on putting in their own postcodes (Figure 9). This normally revealed an area of about 20 houses.

The following data is based on the observations made by the POs at the time. Students were observed to start asking further questions independently very quickly and navigating the map to try to find the answer (see findings for RQ2 in figure 6). For example, one group very quickly put in the postcode for their school. They discovered an anomaly in the data, where a non-building in the school was identified as having good potential for fitting a solar panel. Another group tried to find a building (a head office of a famous pizza chain) that they knew “has a very big roof” to see how much the panels would cost and how much energy it would produce. By querying the data more closely, students latched onto the idea of cost/benefit trade-off. This was despite such activities not being prompted: these students were meant to still be in a *guided* inquiry and there were no open inquiry activities planned for these

Figure 11: Measuring the slope of the roof.

672 younger students.

673 *7.3. Answering RQ3: How does personally collecting data changes one's per-*
 674 *spective of it?*

675 The following results focus on the LP3 activities of FS4, which took place
 676 directly after the LP2 activities described above, where students were explor-
 677 ing the LiDAR data through the technology. The data is based on analysing
 678 and constructing a narrative from the observations of the POs and video data
 679 from the session.

680 There were three groups completing the task, with 2 or 3 children in each
 681 group. All groups completed the task of creating and measuring the house
 682 (figure 11).

683 One observed group were able to complete their grid of height values
 684 taken by measuring the roof height for each square they had drawn onto the
 685 house and then begin to identify the slope of the roof from the data alone.
 686 With some support from the PO, they were working out how they would tell
 687 just from data which way the house was facing (figure 12).

688 Two of three groups swapped their grids and were able to find the slope
 689 from the other group's data, with one group correctly identifying that the

Figure 12: Recreating the house from data

690 other group had made a house with a ‘wiggly roof’ and then asking to see
 691 the house for themselves.

692 At the end of this classroom session, there was a general discussion. The
 693 noted observations were as follows:

- 694 • Students commented how “stupid” the data is, because it “doesn’t
 695 know it is looking at a house, or someone’s back garden”.
- 696 • Students could easily think of things that might have slopes that the
 697 aerial survey might pick up but were not roofs, including bus shelters
 698 or hills.
- 699 • Students thought that it was normally better for humans to process
 700 visual data, but when the data set is so large (as in this case), then it
 701 is good to give some intelligence to computers so they can help.
- 702 • This last comment prompted a discussion about how to add more intel-
 703 ligence to the data processing algorithm. One suggestion was that bus
 704 shelters would not have such a steep pitch. Students started thinking
 705 about combining data sets, proposing that one way to tell a house from

706 other buildings through the data was to measure the heat of people in-
707 side.

708 Taking all of the above into consideration, it appears that the LiDAR task
709 has prompted a good level of understanding of the potential and limitations
710 of the data set (see main findings for RQ3 in figure 6), whilst the initial
711 task with the interactive map prompted more free exploration and asking
712 questions from the data itself.

713 8. Discussion

714 We begin this discussion by considering what has been found with regard
715 to students' abilities to ask and answer questions from externally sourced
716 data.

717 In the fourth field study (FS4), when students were able to directly inter-
718 act with data through the data browser (figure 3), they became very keen to
719 start driving their own inquiries, even though this was not an explicit part
720 of the task (RQ2 in figure 6). For example, deciding to look at the cost of
721 solar panels on a large roof and finding out whether their own houses should
722 get solar panels or not. As pointed out by Konold and Higgins [22], data
723 investigations start with questions about the real world - but such questions
724 must be revised to ones that can be answered from data. The expansion
725 principle (P2) was proposed as a way to support this, by engaging students
726 first with a data snapshot and then allowing them to navigate across the
727 wider data set.

728 In this regard, the finding of note was that younger students (FS3) had
729 more difficulty than older students (FS2) in framing inquiry questions directly
730 from data, when engaging with only a single snapshot (RQ1 in figure 6).
731 Older students were more likely to choose questions for which they could
732 present a plausible explanation of what they would look for in the data to
733 answer. Both sets of students had undertaken an identical task, so the main
734 factor on which to understand the difference was their age. This supports
735 findings of [32] that students of this age find it difficult to link questions, data
736 and explanations coherently. If we take the perspective of Piaget [33], the
737 younger student group are just at the start of their formal operational stage,
738 where they gain the ability to reason in abstract forms. Prior to this stage,
739 children have more reliance on concrete manipulation. If this is the case,
740 then it could explain the observations that the younger students asked more

741 focused questions when they used the technology to engage with the data.
742 However, data collected regarding the role of technology was too sparse to be
743 able to draw firm conclusions and future work would need to investigate more
744 thoroughly the extent to which the technology supported this adaptation of
745 question strategy and played a role in supporting the expansion principle.

746 Turning attention to the personal data collection principle (P6), it was
747 notable that students in the smart meter task (LP1) consistently proposed
748 a supply failure as the reason for a zero reading, whereas a more plausible
749 explanation given the very brief time of the zero reading was that the meter
750 itself had failed. While the results reported have been quite focused, it is fair
751 to mention here that these tasks were conducted across a two year period
752 in a number of settings. It was observed across a number of engagements
753 with smart meter data and also the solar panel data set that students were
754 reluctant to attribute errors to the measuring instrument.

755 In previous work by Hautea et al. [12] it was discovered that young people
756 became sceptical about data through their interactions with it. In this set-
757 up, the students (of a similar age range to the ones in these studies) were
758 interacting with data in an environment in which they were also contributing
759 to the data, so in effect the personal data collection principles was in place to
760 help the students to understand better the possible source of errors. Similarly,
761 in our studies when students started to collect data and became a LiDAR
762 measuring instrument, they were more critical of the data (RQ3 in figure
763 6). These same students had interacted with the LiDAR-obtained solar data
764 in the previous week and had been observed to focus on driving their own
765 inquiries from the data to find if houses were more or less suitable for solar
766 panels. However, in the following week when they were learning how the data
767 was collected, they began questioning whether every ‘roof’ picked up in the
768 dataset was a viable building for fitting solar panels and even started to think
769 of ways to refine the processing of data to reduce such errors. This seems
770 to support the **personal data collection principle** (P6), that students
771 should collect data themselves to help them to interpret data and that this
772 process of interacting with familiar data may be important in fostering data
773 scepticism. In this regard, it would have been better to have these activities
774 occur in the alternate order, so that students would first understand how the
775 data was collected and then explore the data set.

776 The personal data collection principle should be investigated in a more
777 controlled manner, to really understand the relationship between familiarity
778 with data and ability to critique. It has wide-ranging implications for people’s

779 ability to use externally sourced data, whether it is for business needs, for
780 empowerment or for innovation from data.

781 Finally, this work has demonstrated the many different ways that these
782 types of less typical classroom data and smart city concepts can be integrated
783 in a school curriculum and how activities can be designed around them in
784 a way to support development of critical data literacies. Overall, the lesson
785 plans can be shown to achieve their intended outcomes. In the first lesson
786 plan, students showed evidence of finding and explaining common patterns
787 in energy data. In the second lesson plan, students demonstrated a good
788 understanding of the different factors that effect solar yield. In the third
789 lesson plan, students came to understand how to recreate the 3D world from
790 2D data and the possible sources of error that came from the measuring
791 technique. However, this was not the end of the story. Students showed
792 evidence of learning a lot more, for example about the domain of energy, the
793 importance of being energy efficient and the pros and cons of solar energy as
794 a renewable source.

795 9. Conclusions

796 This paper presents findings from an initiative to take complex data from
797 a smart city project into schools and to use it as a teaching resource. It
798 explores the use of data literacy activity design principles to support the
799 co-creation, with teachers, of the teaching resources and the development of
800 technology to support interaction with data. The project followed a research
801 through design approach which created an initial set of teaching materials
802 that were refined each time they were taken to a new classroom and also
803 adapted by the teacher to fit the new context. The technologies to support
804 data interaction were designed to have limited functionality and to support
805 just a small part of the classroom delivery, which also included workbook
806 activities, and practical tasks.

807 The main findings were that:

- 808 • younger students require support in framing inquiry questions that can
809 be answered from externally sourced data;
- 810 • when engaging with externally sourced data it can be useful to act in
811 the role of a data collector to understand better where errors can creep
812 into the data and to develop better data scepticism.

Overall, the learning of data skills lends itself very well to cross-curricular learning and can begin with students as young as ten years old, as evidenced through the variety of school contexts in which we worked. Data literacy activity design principles provide a way to structure learning from external data sets. This may support teachers to develop new activities from open data. The teaching of data in context is important and local, open data can be a good resource for teaching, if supported in the right way.

10. Acknowledgements

This work was undertaken as part of MK:Smart, supported by the Higher Education Funding Council for England [Q-13-037-EM, 201417]. We would like to thank the teachers who welcomed us into their schools and who made this work possible.

11. References

- [1] J. Mateos-Garcia, G. Windsor, S. Roseveare, *Analytic Britain: Securing the right skills for the data-driven economy*, London: Nesta (2015).
- [2] R. Mortier, H. Haddadi, T. Henderson, D. McAuley, J. Crowcroft, *Human-data interaction: The human face of the data-driven society* (2014).
- [3] R. Bhargava, E. Deahl, E. Letouzé, A. Noonan, D. Sangokoya, N. Shoup, *Beyond data literacy: reinventing community engagement and empowerment in the age of data*, Data-Pop Alliance White Paper Series. Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Lad and Overseas Development Institute) and Internews (2015).
- [4] P. Cobb, K. McClain, *Guiding inquiry-based math learning*, Cambridge University Press, 2006.
- [5] M. Janssen, Y. Charalabidis, A. Zuiderwijk, *Benefits, adoption barriers and myths of open data and open government*, *Information systems management* 29 (2012) 258–268.
- [6] K. Kastens, M. Turrin, *Geoscience data puzzles: developing students' ability to make meaning from data*, in: *AGU Fall Meeting Abstracts*.

- 843 [7] L. Bowler, A. Acker, W. Jeng, Y. Chi, “it lives all around us”: As-
844 pects of data literacy in teen’s lives, *Proceedings of the Association for*
845 *Information Science and Technology* 54 (2017) 27–35.
- 846 [8] J. Zimmerman, J. Forlizzi, S. Evenson, Research through design as a
847 method for interaction design research in hci, in: *Proceedings of the*
848 *SIGCHI conference on Human factors in computing systems*, ACM, pp.
849 493–502.
- 850 [9] W. Gaver, What should we expect from research through design?, in:
851 *Proceedings of the SIGCHI conference on human factors in computing*
852 *systems*, ACM, pp. 937–946.
- 853 [10] A. Wolff, D. Gooch, J. J. Cavero Montaner, U. Rashid, G. Kortuem,
854 Creating an understanding of data literacy for a data-driven society,
855 *Journal of Community Informatics* 12 (2017) In–press.
- 856 [11] E. Deahl, Better the data you know: Developing youth data literacy in
857 schools and informal learning environments (2014).
- 858 [12] S. Hautea, S. Dasgupta, B. M. Hill, Youth perspectives on critical
859 data literacies, in: *Proceedings of the 2017 CHI Conference on Human*
860 *Factors in Computing Systems*, ACM, pp. 919–930.
- 861 [13] B. Alper, N. H. Riche, F. Chevalier, J. Boy, M. Sezgin, Visualization lit-
862 eracy at elementary school, in: *Proceedings of the 2017 CHI Conference*
863 *on Human Factors in Computing Systems*, ACM, pp. 5485–5497.
- 864 [14] A. V. Maltese, J. A. Harsh, D. Svetina, Data visualization literacy: In-
865 vestigating data interpretation along the noviceexpert continuum, *Jour-*
866 *nal of College Science Teaching* 45 (2015) 84–90.
- 867 [15] P. Wiegand, Children’s understanding of maps, *International Research*
868 *in Geographical and Environmental Education* 8 (1999) 66–68.
- 869 [16] S. Srikant, V. Aggarwal, Introducing data science to school kids, in:
870 *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Com-*
871 *puter Science Education*, ACM, pp. 561–566.
- 872 [17] V. R. Lee, J. R. Drake, J. L. Thayne, Appropriating quantified self tech-
873 nologies to support elementary statistical teaching and learning, *IEEE*
874 *Transactions on Learning Technologies* 9 (2016) 354–365.

- 875 [18] S. Van Wart, T. S. Parikh, Increasing youth and community agency in
876 gis, in: GeoHCI Workshop at CHI.
- 877 [19] S. Van Wart, K. J. Tsai, T. Parikh, Local ground: A paper-based
878 toolkit for documenting local geo-spatial knowledge, in: Proceedings
879 of the First ACM Symposium on Computing for Development, ACM,
880 p. 11.
- 881 [20] S. Dasgupta, B. M. Hill, Scratch community blocks: Supporting children
882 as data scientists, in: Proceedings of the 2017 CHI Conference on Human
883 Factors in Computing Systems, ACM, pp. 3620–3631.
- 884 [21] T. R. Shelley, C. Dasgupta, A. Silva, L. Lyons, T. Moher, Photomat: A
885 mobile tool for aiding in student construction of research questions and
886 data analysis, *Technology, Knowledge and Learning* 20 (2015) 85–92.
- 887 [22] C. Konold, T. L. Higgins, Reasoning about data, A research companion
888 to principles and standards for school mathematics 193215 (2003).
- 889 [23] S. Williams, E. Deahl, L. Rubel, V. Lim, City digits: Developing
890 socially-grounded data literacy using digital tools, *Journal of Digital
891 Media Literacy* (2014).
- 892 [24] R. Bhargava, R. Kadouaki, E. Bhargava, G. Castro, C. D’Ignazio, Data
893 murals: Using the arts to build data literacy, *The Journal of Community
894 Informatics* 12 (2016).
- 895 [25] C. Anslow, J. Brosz, F. Maurer, M. Boyes, Datathons: an experience
896 report of data hackathons for data science education, in: Proceedings of
897 the 47th ACM Technical Symposium on Computing Science Education,
898 ACM, pp. 615–620.
- 899 [26] C. D’Ignazio, Creative data literacy, *Information Design Journal* 23
900 (2017) 6–18.
- 901 [27] C. Maybee, L. Zilinski, Data informed learning: A next phase data
902 literacy framework for higher education, in: Proceedings of the 78th
903 ASIS&T Annual Meeting: Information Science with Impact: Research
904 in and for the Community, American Society for Information Science,
905 p. 108.

- 906 [28] R. Bhargava, C. D'Ignazio, Designing tools and activities for data liter-
907 acy learners, in: Workshop on Data Literacy, Webscience.
- 908 [29] A. Wolff, A.-M. Valdez, M. Barker, S. Potter, D. Gooch, E. Giles,
909 J. Miles, Engaging with the smart city through urban data games,
910 in: Playable Cities, Springer, 2017, pp. 47–66.
- 911 [30] A. Wolff, J. J. C. Montaner, G. Kortuem, Urban data in the primary
912 classroom: bringing data literacy to the uk curriculum, *The Journal of*
913 *Community Informatics* 12 (2016).
- 914 [31] B. B. Kawulich, Participant observation as a data collection method, in:
915 *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*,
916 volume 6.
- 917 [32] H.-K. Wu, C.-E. Hsieh, Developing sixth graders inquiry skills to con-
918 struct explanations in inquiry-based learning environments, *Interna-*
919 *tional Journal of Science Education* 28 (2006) 1289–1313.
- 920 [33] J. Piaget, B. Inhelder, *The growth of logical thinking from childhood to*
921 *adolescence: An essay on the construction of formal operational struc-*
922 *tures*, Routledge, 2013.