



Experimenting with Dataphys, a Physicalization Toolkit for Representing Spatio-Temporal Data

Fabio Pittarello
pitt@unive.it

Università Ca' Foscari Venezia
Venezia, Italia

Manuel Semenzato
857797@stud.unive.it

Università Ca' Foscari Venezia
Venezia, Italia

ABSTRACT

Data physicalization offers an exciting opportunity for engaging people with physical representations of data that can be examined and manipulated to gain a mental model of a set of variables and their relationships. This is of paramount importance not only for researchers and professionals but also for improving citizens' awareness in different domains, impacting civil society, and pushing behavior change. Accessibility of data is another important issue that can benefit from the adoption of data physicalization. Despite this potential, several pitfalls prevent the wide adoption of this paradigm, including the lack of a methodology for guiding the design of the representations and toolkits permitting to engage citizens in their creation without requiring them to start from scratch. This paper intends to contribute to the discussion by presenting and evaluating Dataphys, a toolkit for creating physical representations of data. The toolkit is modular, conceived for the application to different scenarios and domains, printed with cheap technologies, and open to modifications to ease its adoption. The evaluation was carried out in the context of a workshop with the participation of 17 master students in computer science, and focused on the gathering of environmental data collected along a path and the construction of a physical representation. The study's results confirm the interest in the approach and offer insights for improving it.

CCS CONCEPTS

• **Human-centered computing** → **Visualization design and evaluation methods; Empirical studies in visualization; Visualization toolkits; Interface design prototyping**; • **Social and professional topics** → *Sustainability*.

KEYWORDS

data physicalization, environmental data, evaluation, methodology, toolkit

ACM Reference Format:

Fabio Pittarello and Manuel Semenzato. 2024. Experimenting with Dataphys, a Physicalization Toolkit for Representing Spatio-Temporal Data. In *International Conference on Advanced Visual Interfaces 2024 (AVI 2024)*, June 03–07, 2024, Arenzano, Genoa, Italy. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3656650.3656682>



This work is licensed under a Creative Commons Attribution-Share Alike International 4.0 License.

AVI 2024, June 03–07, 2024, Arenzano, Genoa, Italy
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1764-2/24/06
<https://doi.org/10.1145/3656650.3656682>

1 INTRODUCTION

Access to data and the possibility of forming a mental model starting from a proper representation of them is becoming increasingly important nowadays, not only for researchers and professionals but also for common citizens, for improving their awareness in different domains and pushing behavior change. Data can be communicated in different fashions, and information visualization provides a variety of techniques to present data for different devices and contexts. Engaging citizens with representations is of paramount importance because it grants attention and is unanimously recognized as a prerequisite for learning [11]. Tangible representations have a potential in this respect because they present information as objects that belong to the real world, whose components can be manipulated during the creation phase and often during the presentation to the final users. Besides, tangible representations also have the potential for accessibility because they can provide a way for non-sighted users to form a mental model of a given domain.

However, while there are examples of physical representations of data, a toolkit based on a design methodology taking into consideration all the steps of the physicalization process is still missing. This paper intends to contribute to this research by evaluating a toolkit supporting the users in forming a mental model of a given domain. The toolkit is based on a methodology presented in [21], that was accompanied by a version of the toolkit that missed an evaluation study. The evaluation presented in this work was carried out in a workshop focused on gathering and representing environmental data, which represents a relevant application domain. Making citizens aware of issues related to the environment becomes more and more urgent to involve them in behavioral changes. The representation of data is an important activity, but involving citizens in collecting data themselves also has an important role in convincing people to trust information. For this reason, the workshop included an initial data gathering phase. The evaluation is meant to answer the following research questions related to the toolkit:

- Q1. Did the physicalization toolkit impact the formation of a mental model about a given domain?
- Q2. Were the toolkit components properly designed to represent variables and spatio-temporal context?
- Q3. Which were the points of strength and weakness and possible improvements?
- Q4. Can results be extended to other application domains?

Besides, the workshop was also useful for receiving feedback on the impact of data gathering and its impact on data understanding.

2 RELATED WORKS

According to Kirk [17], data visualization is *the visual representation and presentation of data to facilitate understanding*. The underlying

idea of information visualization is that if we move from a simple description of raw data (e.g., a tabular description) to a proper representation, we can obtain insights that can't be perceived in the initial description, understanding more about the variables involved and their relations and ask ourselves new questions about what stems from the representation [25]. While Kirk's definition focuses on visual representation, and many examples take advantage of the visual channel for delivering information, other researchers have focused on the proposal of representations targeted to other senses. Data sonification [31] takes advantage of audio representations to convey information. The sense of smell is less used because of the difficulty of digitizing and reproducing a variety of fragrances, but, in recent years, some interesting experimentations have been carried out [5].

The interest in the use of sense of touch and the creation of tangible representations has rapidly grown in the last years, leading to the definition of a research area named *data physicalization*. According to Jansen et al. [15], data physicalization is *a research area that examines how computer-supported, physical representations of data can support cognition, communication, learning, problem-solving, and decision-making*. A physicalization is defined as *a physical artifact whose geometry or material properties encode data*. While many academic works have explored its potential in recent years, the first examples of this type of representation trace back to clay tokens used in Mesopotamia to represent units of measurement for different goods [24]. The site *dataphys.org* [8] shows a comprehensive timeline of physical data visualization throughout history. In modern times, before the definition of data physicalization, Zaho et al. [32] have introduced the concept of *data sculpture* defined as *data-based physical artifact possessing both artistic and functional qualities, that aims to augment a nearby audience's understanding of data insights and any socially relevant issues that underlie it*.

In recent work, Bae et al. [2] summarize the benefits of data physicalization, mentioning the broadening of participation, the support for analytics, and the promotion of creative expression. Some challenges, such as those related to fabrication and interpretation [3], are also highlighted. Other recent works highlight the benefits of data physicalization for increasing engagement and participation in educational scenarios [20] [26] and promote awareness of complex data in a non-expert audience [18] [22] [23].

The audience's involvement in creating physical artifacts representing data makes information more accessible, understandable, and democratic. This is an approach, named *constructive visualization*, suggested by Huron et al. [14] that define it as *the act of constructing a visualization by assembling blocks that have been previously assigned a data unit through a mapping*. The creation process can start from raw materials [20] [26], but some authors also proposed the use of ready-made tokens [1] [23].

Having a methodology for mapping data and their relations to the visual, auditorial, physical, or olfactory properties that can be perceived and distinguished by human beings is a fundamental step for creating successful representations. Jacques Bertin was a pioneer in this respect and in 1967 identified six *retinal variables* (i.e., color hue, color value, size, shape, orientation and texture) that can be used in the creation of visual representations of entities to ease tasks such as their selection, association, ordering, and identification of quantities [4]. The proposal was presented as a

diagram that had a significant influence on the following discussion, even if alternative mappings were proposed later (e.g., Cleveland et al. [7] in 1984). However, the main focus of these proposals was on visual representations. More recently, Stusak et al. [27] mapped the same four tasks identified by Bertin to a set of physical variables that took into consideration geometric, color, tactile, and kinesthetic properties of physical objects (see Table 1).

In a relevant number of situations, data are collected in relation to specific locations and then translated into physicalizations. It is useful to classify three different spatial situations:

- *point*: data are collected in a single location for a time span, as in *Summer in the City* [30] where data of air pollution were collected during 4 weeks in the summer of 2015;
- *grid*: data are collected, simultaneously or not, in different locations that can also be discretized; *Global Cities* [29] is an elevation map representing the population density of 12 of the world's major urban centers;
- *path*: data are collected along a path, usually by a single gathering device.

Data walking is a research project related to the third spatial scenario. Hunter [12] designed this project proposing a series of workshops in different locations worldwide. In these workshops, participants were involved in creating devices for gathering environmental data, collecting data by walking, creating the representations, and speculating on the results. Different data representations were created, including visual, auditorial, and physical artifacts [13]. *Data cylinders* is one of the most interesting physical representations described in Hunter's work. Walks are represented through a stack of cylindrical washers whose size maps the values of an environmental variable (e.g., light, sound, temperature, or air quality). Besides, each surface of the cylinder embeds a picture of where the data was collected.

Finally, while data physicalization initially didn't focus on accessibility [6], it has a potential for improving inclusiveness, with a special reference to non-sighted persons. While there is a lot of work for the conversion of hypertextual information into an accessible format (e.g., W3C Web Accessibility Initiative (WAI) and the related accessibility guidelines [28]), less work has been done to design approaches for converting structured data in an accessible format. Some approaches try to fill in the gap by proposing heuristics [9] or tools for automatically converting charts already available on the web into textual data tables [6], that then can be accessed with a text-to-speech engine or a braille keyboard. While the latter approach is very interesting for permitting basic access to data, it transforms data in tables, which often are not useful for permitting users to create a mental model of data themselves [25]. Mapping data to physical objects can offer a better opportunity in this respect, even if it should be done properly and using accurate technology, as underlined by Lundgard et al. [19]. Again, authors such as Kane et al. [16] outline the importance of involving users in creating physical representations, as they did in a workshop with non-sighted high-school students.

Concluding, the analysis of related literature shows that while many interesting studies related to physicalization are available, a systematic methodology guiding the designer to consider all the necessary steps for designing a toolkit that eases the creation of

physical representations, is missing. Reusability is an important related issue. While the design of a component-based toolkit can somehow limit the expressivity, it may permit the creation of meaningful and aesthetically pleasing representations without starting from scratch. It can make a difference in many situations where limited resources and custom designs for mapping data to physical representations can't be afforded.

3 THE METHODOLOGY

This section summarizes the methodological approach described in [21], meant to support the designers in considering all the necessary steps to define a physicalization toolkit.

3.1 Mapping data to physical variables

The mapping of the physical variables to the four tasks associated to data understanding, described by Stusak et al. [27], represents the first step of the proposed methodology. Yet, this mapping doesn't consider all the constraints brought by different human, contextual, and technological factors that can influence the design of a toolkit. The following subsections will consider the impact of the different factors. For clarity, each step will be accompanied by selecting a specific context, technology, and human profile that will determine the specific features of the toolkit. Of course, different choices would lead to different constraints and different solutions for the final toolkit. Yet, the toolkits resulting from this approach are general enough to be used in different types of contexts and domains without requiring to restart from scratch for each specific application.

3.2 The role of technology

The technology selected for the production of the toolkit has a relevant impact on the final result. While different technologies are available nowadays, there are relevant differences in terms of expressivity and costs. Because one of the goals of this work is to make access to information more democratic, in compliance with Huron et al.[14], we selected the widely diffused FDM printing technology, which works by extruding a thermoplastic filament. Several constraints characterize this technology, including the limited printed volume, the resolution and the smoothness of the print, the printing time, and the very limited number of colors that can be used simultaneously. To maximize the diffusion of the proposal, we decided also to refer to a basic printer that uses only a single filament at once. This led us to introduce some limitations in the use of the physical variables introduced by Stusak et al. that can be summarized as follows:

- A: only a single value of the variable can be used (i.e., *coldness* has a single value because the thermal conductivity of all the plastic filaments used by FDM printers is the same);
- B: only a limited set of values of the variable can be used; i.e., only a limited range of values for *hue*, *saturation*, *luminance*, *compliance* and *slipperiness* can be obtained with commercially available materials;
- C: the values that can be used for a given variable depend on the values of the other variables (i.e., it may be possible to obtain only specific combinations of values for *hue*, *saturation*, *luminance*, *optics*, *compliance* and *slipperiness*).

Table 1: The physical variables by Stusak et al. mapped to the four tasks (*X* stands for good mapping, *o* for possible mapping). Column *Lim.* identifies the different types of limitations: *A* (only a single value of the variable can be printed), *B* (only a limited set of values of the variable can be printed), *C* (the values that can be printed for the variable depend from the values of other variables).

Variables	Selection	Associat.	Quant.	Order	Lim.
<i>Geometric</i>					
Position	X	X	X	X	
Orientation	X	X		o	
Global Shape	X	X	o	X	
Exact Shape	X	X		o	
<i>Color</i>					
Hue	X	X			B,C
Saturation	X	X		X	B,C
Luminance	X	X		X	B,C
Optics	X	X		X	C
<i>Tactile</i>					
Roughness	X	X		X	
Lay	X	X		o	
Coldness	X	X		X	A
Compliance	X	X		X	B,C
<i>Kinesthetic</i>					
Slipperiness	X	X		X	B,C
Weight	X	X		X	

All these limitations are reported in the last column of Table 1.

3.3 The role of the spatial context

Data may belong to different spatial contexts and this can impact their mapping to a physical representation. In Section 2, three types of situations have been described related to gathering data in a single location, grid of locations, or paths. The temporal dimension associated with the spatial context (i.e., data collected in parallel or sequence) can have an additional impact. The toolkit proposed in this work will focus on a specific spatio-temporal scenario related to data collected in a sequence along a path, which is common to several situations related to data gathering. To come to the final solution, we analyzed *Data cylinders*, one of the most interesting solutions available in the literature, because of its simplicity, attractiveness, and suitability to represent different types of data along a path. We share with the *Data cylinders* approach the interest for a simplified approach that considers only the temporal sequence of data gathering and the possibility of manipulating the different slices of the sequence even after the creation. We tried however to improve a number of dimensions for coming to the final proposal, including its expressivity (augmenting the number of variables that can be represented), its clarity (trying to limit occlusion issues and giving up to the pictures of the locations that were not identifiable in the stack) and the support for manipulating the stack.

3.4 The role of humans

Human beings have potential and limits that should always be considered when designing a representation targeted to them. Our methodological approach intends to make explicit that the mapping proposed by Stusak et al. is meant for able-bodied users and that additional requirements must be considered for different classes of users. For example, for non-sighted users, all the variables related to the *Color* class (i.e., hue, saturation, luminance, optics) can't be considered for coming to the final solutions. While the toolkit evaluated in this work is targeted to able-bodied, in [21] we describe how the specification of constraints for non-sighted users can lead to designing a different and more accessible toolkit.

4 THE PHYSICALIZATION TOOLKIT

The physicalization toolkit described in this section stems from the application of the aforementioned methodology after the application of specific design requirements (FDM printing, spatio-temporal sequence, normal-bodied users). The toolkit design represents an important addition to the methodology that enables users to create physical representations in various situations without starting from scratch. Flexibility, modularization, and expandability are the main features of the toolkit, meant to ease its application with different data and in different domains. The toolkit includes the following components:

- *tokens*: cylindrical shapes printed in a single color and available in five different sizes; they are meant to represent the main variable and, when composed in a stack, to display the spatio-temporal sequence of the gathering path; each cylinder is provided with eight holes for plugging additional components, and magnets to ease the cylinders' connections (see Fig.1); the height of the cylinders was arranged for diminishing occlusion problems;
- *add-ons*: small components printed in different colors and shapes that can be plugged into the cylinders for mapping the values of additional variables.

The design of the cylinder-based structure and its composition in a vertical tower resulted from a trade-off that hid the full-path representation, limiting the information to the spatio-temporal sequence. A *layer* of the representation stems from the composition of a single token with one or more add-ons, and it is referred to a specific location and time. Add-ons can be split into two different categories:

- *value add-ons*: they are available in different shapes, sizes, and colors and are designed to qualify a set of values of a given variable (see Fig.2 B, C, D and E);
- *markers*: they are available in a single shape, size, and color and are designed to qualify the presence of a given variable (see Fig.2 F)

During the design phase, we debated, with contrasting arguments, the relevance of representing data with icons or abstract shapes. In the end, we decided to provide both *abstract* shapes (e.g., circles and pins) and *iconic* shapes (e.g., houses, trees, and magnifying glasses) and test them on the field.

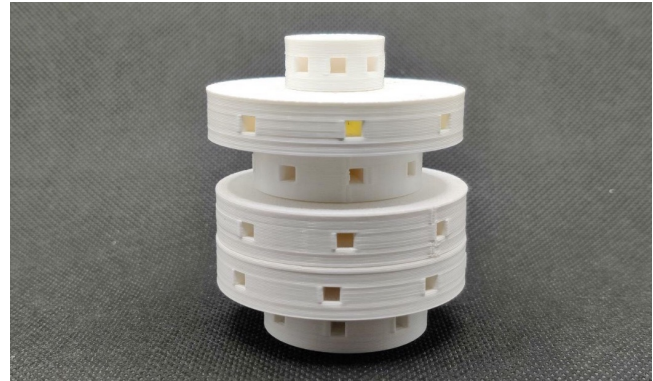


Figure 1: Tokens composed in a vertical tower; each cylinder has eight holes and contains magnets for easing the connections.

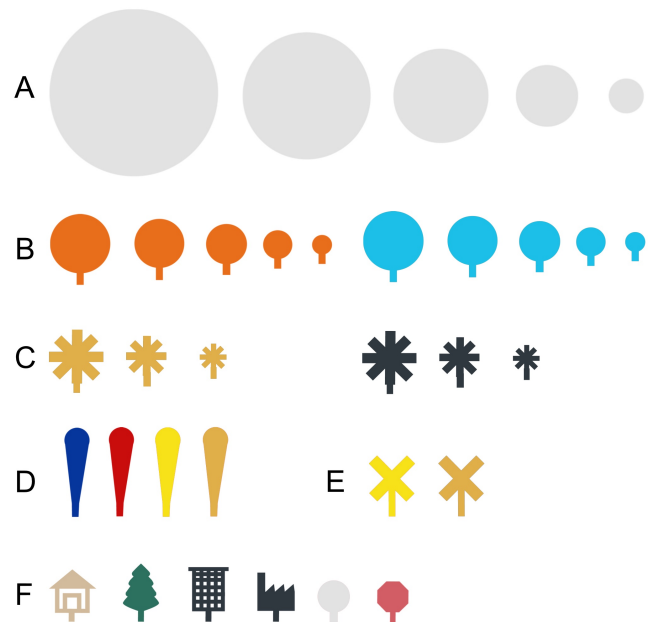


Figure 2: Visual abacus of tokens (A), value add-ons (B, C, D and E) and markers (F).

5 CASE STUDY: GATHERING AND REPRESENTING ENVIRONMENTAL DATA ALONG A PATH

To test the points of strength and weakness of the toolkit and its suitability to represent data collected along a path, we organized a workshop focused on the gathering of carbon dioxide and other environmental data around the University Campus and the creation of physical representations based on the data collected. As underlined in the Introduction, environmental data represents a relevant application field. Citizens' science [10] initiatives show the importance of involving people in activities related to environmental data gathering to improve their awareness and push behavior

changes. Initiatives like *Data walking* [12] show the additional benefits of involving people in creating representations stemming from these data. To lower the barrier for participants, volunteers didn't need to create their own devices for gathering data, as in the *Data walk* workshops, but were given pre-assembled portable stations (see Fig.3). While the full discussion of these devices is out of the scope of this paper, we'll mention that these network-connected devices allowed to collect carbon dioxide concentration and other environmental data (humidity and temperature), taking advantage of a simple interface for activating the gathering and monitoring their values in real-time.

5.1 Workshop structure

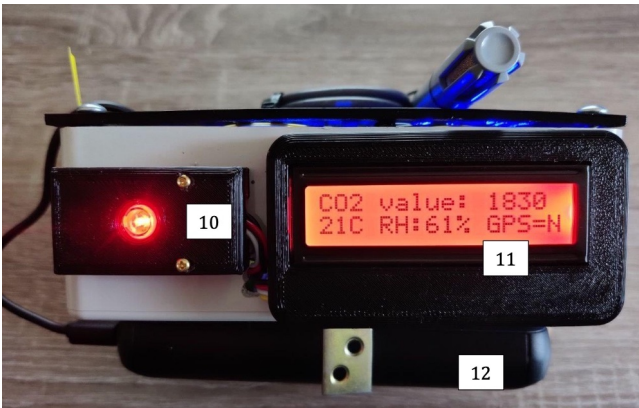


Figure 3: The mobile station for gathering carbon dioxide and other environmental data.



Figure 4: Three different contexts for the walking experience: urban, green, and industrial environments.

17 master students in computer science took part in the workshop that was organized as follows:

- *initial briefing*: the students were given a brief introduction about the goal of the workshop (creating a physicalization of a set of environmental variables collected by walking along an assigned path) and some background information about carbon dioxide, the main variable to be collected, and how it relates to climate change issues. Participants, with the only exception of a group that was composed of three persons to cope with the odd number of volunteers, were grouped in pairs and assigned a path to walk bringing with them the portable station. The length of each path was around 2 km, and each of them had the University Campus as the final destination. Three different types of contexts were explored,



Figure 5: Creating the physical model with the toolkit.

as outlined by the pictures in Fig.4: an urban environment, a green environment and an industrial environment;

- *data gathering*: each pair of participants received a mobile station and a link to a digital map, accessible from their smartphones, that included the precise path to be walked. Each path was divided into 6-7 sections and participants were invited to annotate in a blank paper template additional qualitative environmental information (i.e., type of environment, landmarks, wind speed, environmental noise, air smell, quantity of cars and quantity of people). Participants were left free to select how many levels to use for each qualitative variable (e.g., strong wind, weak wind, no wind);
- *data pre-processing*: after the gathering, data collected through the portable stations were pre-processed by the team that organized the workshop for determining the means of the three variables automatically gathered (carbon dioxide, temperature, humidity). These means were then made available to the workshop participants;
- *creation of the physicalizations*: the participants were invited to create the physicalization for their path, starting from the data collected with the devices and observed (an example of the result can be seen in Fig.6 and Fig.7). Each group was given a copy of the toolkit (see Fig.2) and worked independently (see Fig.5). All the participants were asked to use the cylindrical shapes for mapping the carbon dioxide and, for easing comparisons, we agreed on the size of the cylinder to associate with a given interval of the CO₂ concentration. For what concerns the qualitative variables we asked the groups to consider only a subset of them (i.e., type of environment, landmarks, air smell and quantity of cars), to understand more clearly their preferences for the different types of toolkit add-ons;
- *survey fill-in*: at the end of the workshop, the participants were asked to fill in a survey, which included both closed questions based mainly on a 5-point Likert scale and some open questions. The survey included the following sections:

- *demographics*, including: demographic information and involvement in previous experiences related to environmental data gathering;
- *feedback related to the experience of data gathering*, including: information related to the cognitive and physical load of the gathering experience; the possible improved awareness of the environment given by the gathering experience; the impact of the gathering on data understanding;
- *feedback related to the experience of creating the physical representations*, including: the possible improved engagement compared to visual representations; the preference for abstract and iconic shapes for representing data; the usefulness of the physical shapes for representing environmental data or other data types; the mapping of geographical features; the impact of creating the physical representation for data understanding; finally a set of open questions for receiving feedback about the points of strength and weakness of the approach and suggested improvements.

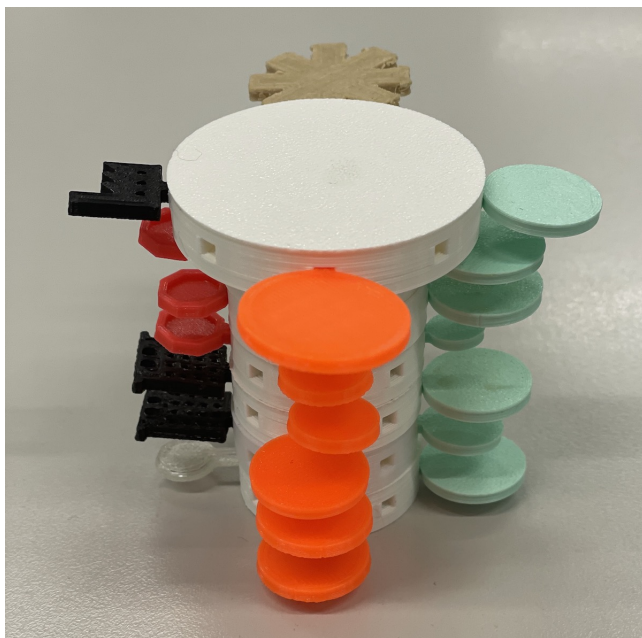


Figure 6: One of the physical representations realized by the students, mapping carbon dioxide data and other four environmental variables.

5.2 Results

All the 17 students participating in the workshop were attending a master in computer science. There were 13 males and four females, with a mean age of 24.2. 70% of them didn't have any previous experience in gathering environmental data. Concerning the data-gathering experience, Fig.8 shows that, for most students, the walk was not demanding from a cognitive and physical point of view. Most students stated that, because of the workshop involvement, they paid attention to environmental features they usually didn't notice during the walk. Therefore, the declared focused attention

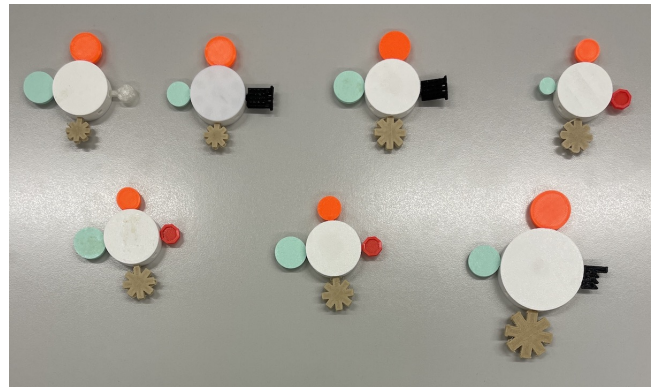


Figure 7: Another view of the same physical representation, with the single cylinders disposed on a table.

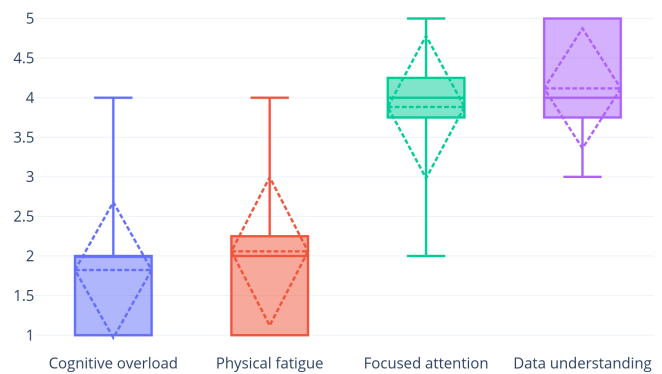


Figure 8: Cognitive load, physical fatigue, focused attention, and data understanding during the data gathering phase. Means and SDs are reported using dashed lines.

was very high (mean 3.9 sd 0.90). Besides, most students agreed that the experience of gathering data in person had positively impacted their understanding of data (mean 4.1 sd 0.76).

Concerning the following phase, related to the creation of the physical representation, the survey gave helpful feedback that we can classify into several areas: engagement, data representation, spatial representation, and impact on data understanding.

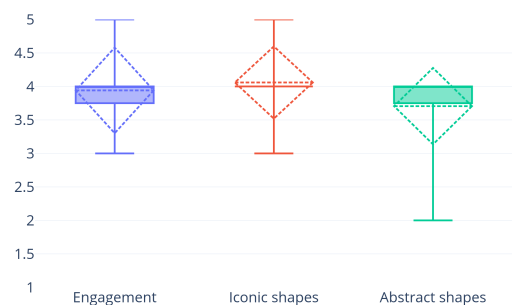


Figure 9: Engagement provided by the physical representation of data - Perceived value of iconic and abstract add-ons.



Figure 10: Perceived usefulness of the toolkit for representing data collected along a path, for the environmental domain and other domains - Interest for the representation of position.

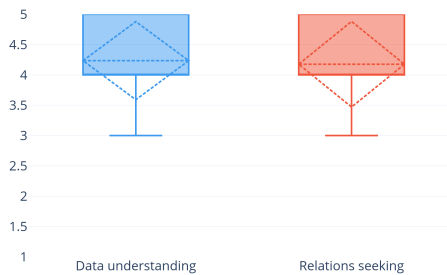


Figure 11: Impact of the creation activity on data understanding and seeking of relations among variables.

The volunteers were asked how engaging they considered the tangible representation of data, compared to a visual representation, and the answers (see Fig.9) display a high level of engagement (mean 3.9 sd 0.64 and no answer below 3 points). The users' answers related to the use of iconic or abstract shapes display an appreciation for both solutions, even if the iconic shapes obtained higher scores (mean 4.1 sd 0.54) and no score below 3 points.

We were eager to hear from the users about the design of the cylinder-based structure, a trade-off that limited the communicated spatial information. The answers (Fig.10) show that they considered the solution appropriate for representing environmental data gathered along a path; the solution was also considered attractive for other domains, although slightly lower scores were assigned. However, the possibility of modifying the toolkit to represent also the geographical position was considered worth consideration (Fig.10). An associated open question investigated possible ideas from the students about how to come to this solution. The most common proposal (4 students out of 17) was about placing the cylinders on the top of a map, positioning each of them in the different sectors of the walking path. Other less intuitive proposals focused on adding a new add-on for each cylinder, indicating the location with a label.

Finally, the question related to the impact of the activity of creating the representation on data understanding (see Fig.11) obtained very high scores (mean 4.2 sd 0.64 and no score below 3). Similar results were achieved about the stimulus provided by the physical representation to find relations between carbon dioxide levels

and the other environmental variables gathered during the walk. An associated open question gives more insights into which environmental variables students tried to relate to carbon dioxide. Many students focused on the number of cars (16 out of 17) or the type of environment (nine out of 17); a smaller number of students considered relations with the number of persons, wind, and air smell.

The final open questions related to the points of strength and weakness and possible improvements gave the following results. Concerning the points of strength, *expressivity* obtained the best results (cited by six students), followed by *simplicity and effectiveness* (five students). Three students mentioned the possibility of *understanding relationships*, while the *identification of the spatio-temporal order* was outlined by two students. Concerning the point of weakness, the main complaint was about the *quality of fabrication* (cited by four students) that sometimes made it difficult to plug the add-ons. Two students mentioned the *partial occlusion* of the stacked representation for certain carbon dioxide patterns. Low expressivity was mentioned only by a couple of students.

Finally, concerning toolkit improvements, most comments were focused on improving further the toolkit expressivity by providing more iconic add-ons (10 students out of 17). A few comments focused on improving the portability of the toolkit and diminishing the partial occlusion by using transparent tokens.

6 DISCUSSION

Concerning the impact of the physicalization toolkit on the formation of a mental model (Q1), the feedback received allows us to give a positive answer to the first request question. The physical representations were perceived as engaging (Fig.9), and this represents a good result because, as already stated in the Introduction, engagement is often a pre-requisite for learning [11]. In addition to the widely recognized role of visual representations over raw data [25], students outlined the impact of the construction activity on data understanding (Fig.11). In this respect, while a possible bias might have come from the involvement of the students in the data gathering, we believe that during this activity students didn't have the time to focus also on the relations between variables. On the contrary, the following manipulation of the toolkit components, in a calm situation, pushed the students to seek possible relations between variables, and a couple of students also claimed some possible findings. We also underline that students' perception is compliant with what is suggested by Piaget about the role of manipulation for learning by children, and by Chapman that extended the results of Piaget's studies to people of all ages [14].

We got positive feedback on the proper design of the toolkit components for representing variables and spatio-temporal context (Q2). The main concept of the physical toolkit, based on tokens for representing the primary variable and indicating the spatio-temporal sequence and the set of add-ons for specifying ancillary variables, was appreciated. Proposals for improvements took advantage of the modular structure of the toolkit and were mainly targeted to suggest additional types of add-ons. The answers to the perceived value of the iconic and abstract shapes (Fig.9) show that it was important to provide both of them to improve flexibility and make the application to different contexts and domains. Students

preferred to use iconic shapes when there was a direct semantic relationship with the variables to map, and the requests for extensions were mainly targeted to new icons. However, we must point out that additional types of add-ons should also be considered in terms of sustainability (i.e., costs and possible waste of materials). The solution probably stands in the availability of a larger number of 3D digital models to print selectively, according to the need of the specific domain. The representation of the spatio-temporal context offered by the toolkit represents a compromise between expressivity and simplification that the students well accepted. However, some recognized the value of providing more information about the path, such as combining the layers with a map.

Concerning the points of strength and weakness and possible improvements of the toolkit (Q3), the answers outline the features that guided the toolkit's design (expressivity, simplicity, data, and context understanding) and, therefore, represent a confirmation of the initial choices. Points of weakness were mainly related to the quality of fabrication, which will be considered for the next release of the toolkit. Complaints about the partial occlusion of some patterns will also be considered. The definition of the cylinders' height was a result of a trade-off between the minimization of occlusion and the need to avoid excessively tall stacks. However, we'll experiment with some design solutions, including spacers for the cylinders that could be used in specific situations.

Concerning the extension of the toolkit to other application domains (Q4), we tried to investigate how the students perceived its usefulness in other contexts (i.e., we suggested cultural heritage as an alternative application domain). The feedback we obtained (Fig.10) was good but not as good as the one obtained for the domain they were directly experimenting with (environmental data). The hypothesis is that the general structure of the toolkit and its modularity led to positive feedback. Still, the lack of expertise in the suggested domain and the lack of specific iconic add-ons led to a less enthusiastic appreciation. This hypothesis is compliant with the fact that the main request for improving the toolkit was related to having more iconic add-ons. Resuming, this issue deserves further investigation, with the involvement of communities directly involved in the application domains under exam.

Overall, the feedback collected in the workshop positively answers all four research questions formulated in the Introduction. Besides, the answers related to data gathering (Fig.8) confirmed the value of participating in these experiences. Students didn't consider the gathering as cognitively or physically overloading, but they paid attention to several details that they perceived as important for understanding the data they gathered. In this respect, the workshop represents a confirmation of the importance of involving users in the gathering of data.

The study was also very useful for examining the different approaches used by the students to create the representations. For example, the students that created the representation displayed in Fig.6 and Fig.7 positioned the add-ons at a regular distance and composed the stack of cylinders, creating virtual vertical columns of add-ons (e.g., the columns of orange and light blue circles or the column of the urban landmarks). This might seem obvious, but other groups operated different choices, using the add-ons to fill in the holes available on one side of the cylinder, placing the add-ons in a different order for each cylinder, or even using more copies

of the same add-on for representing different quantities of a given variable on the same layer. While maintaining the same pattern for each cylinder, other groups didn't bother to compose the cylinders creating the virtual vertical columns displayed in Fig.6. All these practices represent points of weakness in both the comprehension and the accessibility of the physical representation and therefore suggest that providing only the toolkit is insufficient and a set of guidelines for supporting the users in the composition of the representations is needed. These guidelines should be inserted as the last step of the methodological approach. An initial set of guidelines could be resumed as follows:

- compose the vertical stack of cylinders aligning the cylinders' holes to create eight virtual columns; for each ancillary variable, select one of these columns for placing the related add-ons; as a result, each cylinder will have the add-ons related to each variable in the same order; this guideline would be particularly beneficial for non-sighted users that could benefit from a simplified tactile scan of the physical representation [21];
- don't place multiple add-ons for a single variable in the same cylinder; when in need of representing different quantities, use different sizes and colors of the same add-on instead;
- for a given cylinder, to minimize horizontal occlusion, place the add-ons at a regular distance (e.g., in the case of four complementary variables, alternate plugs and free holes).

7 CONCLUSION

The main contribution of this paper is the evaluation of a toolkit for the physicalization of data, characterized by a flexible and modular approach based on a novel methodology that considers different factors for guiding the design of the toolkit itself. Despite some limitations of the study, like the lack of a comparison with a baseline situation where the participants don't have to manipulate the toolkit components but only observe them, the evaluation gave positive results regarding the structure of the toolkit, its impact on data understanding, and its applicability to different domains. Although general enough to be applied to different situations, the proposed toolkit represents a solution for a specific technological and spatio-temporal scenario and human profile. Different design solutions are possible when one or more factors change, as shown in the toolkit for non-sighted users described in [21]. The availability of the methodological approach described in this paper is meant to ease the design of toolkits related to other alternative scenarios. For the same reason, we plan to make the toolkit presented in this work available through a web repository. The goal is to foster the diffusion and democratization of physical representations, supporting extensions, adaptations and proposals of alternative toolkits that could be created starting from our work.

ACKNOWLEDGMENTS

We warmly acknowledge Fabio Pranovi for the insights related to environmental data and the support for the development of the mobile stations. Many thanks to the master students of Ca' Foscari University of Venice, for their participation to the workshop and the creation of the nice graphics that were used for the visual abacus.

REFERENCES

- [1] Lorenzo Ambrosini and Miriah Meyer. 2022. Data Bricks Space Mission: Teaching Kids about Data with Physicalization. In *2022 IEEE Workshop on Visualization for Social Good (VIS4Good)*. 10–14. <https://doi.org/10.1109/VIS4Good57762.2022.00007>
- [2] S Sandra Bae, Danielle Albers Szafrir, and Ellen Yi-Luen Do. 2022. Exploring the Benefits and Challenges of Data Physicalization. In *Proceedings of ETIS 2022*.
- [3] S Sandra Bae, Clement Zheng, Mary Etta West, Ellen Yi-Luen Do, Samuel Huron, and Danielle Albers Szafrir. 2022. Making data tangible: A cross-disciplinary design space for data physicalization. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [4] Jacques Bertin. 1983. *Semiology of graphics*. University of Wisconsin press.
- [5] Anna R. L. Carter, Marianna Obrist, Christopher Dawes, Alan Dix, Jennifer Pearson, Matt Jones, Dimitrios Zampelis, and Ceylan Beşevli. 2023. Scent InContext: Design and Development around Smell in Public and Private Spaces. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference (Pittsburgh, PA, USA) (DIS '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 138–141. <https://doi.org/10.1145/3563703.3591455>
- [6] Jinho Choi, Sanghun Jung, Deok Gun Park, Jaegul Choo, and Niklas Elmquist. 2019. Visualizing for the Non-Visual: Enabling the Visually Impaired to Use Visualization. *Computer Graphics Forum* (2019). <https://doi.org/10.1111/cgf.13686>
- [7] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554. <https://doi.org/10.1080/01621459.1984.10478080> arXiv:<https://www.tandfonline.com/doi/pdf/10.1080/01621459.1984.10478080>
- [8] DataPhys 2023. Data Physicalization Wiki. <http://dataphys.org>
- [9] Frank Elavsky, Cynthia Bennett, and Dominik Moritz. 2022. How accessible is my visualization? Evaluating visualization accessibility with Chartability. *Computer Graphics Forum* 41, 3 (2022), 57–70. <https://doi.org/10.1111/cgf.14522>
- [10] Mordechai (Muki) Haklay, Daniel Dörler, Florian Heigl, Marina Manzoni, Susanne Hecker, and Katrin Vohland. 2021. *What Is Citizen Science? The Challenges of Definition*. Springer International Publishing, Cham, 13–33. https://doi.org/10.1007/978-3-030-58278-4_2
- [11] Juho Hamari, David J. Shernoff, Elizabeth Rowe, Brianno Coller, Jodi Asbell-Clarke, and Teon Edwards. 2016. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior* 54 (2016), 170–179. <https://doi.org/10.1016/j.chb.2015.07.045>
- [12] David Hunter. 2020. Data walking. In *Proceedings of the 16th Participatory Design Conference 2020-Participation (s) Otherwise-Volume 2*. 188–191.
- [13] David Hunter. 2022. *Data Walking*. Retrieved January 21, 2024 from <https://www.datawalking.com>
- [14] Samuel Huron, Sheelagh Carpendale, Alice Thudt, Anthony Tang, and Michael Mauerer. 2014. Constructive visualization. In *Proceedings of the 2014 conference on Designing interactive systems*. 433–442.
- [15] 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*. 3227–3236.
- [16] Shaun K. Kane and Jeffrey P. Bigham. 2014. Tracking @stemxcomet: Teaching Programming to Blind Students via 3D Printing, Crisis Management, and Twitter. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education (Atlanta, Georgia, USA) (SIGCSE '14)*. Association for Computing Machinery, New York, NY, USA, 247–252. <https://doi.org/10.1145/2538862.2538975>
- [17] Andy Kirk. 2019. Data visualisation: A handbook for data driven design. *Data Visualisation* (2019), 1–328.
- [18] Martin Valdemar Anker Lindrup, Arjun Rajendran Menon, and Aksel Bjørn-Hansen. 2023. Carbon Scales: Collective Sense-Making of Carbon Emissions from Food Production through Physical Data Representation. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (Pittsburgh, PA, USA) (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1515–1530. <https://doi.org/10.1145/3563657.3596043>
- [19] Alan Lundgard, Crystal Lee, and Arvind Satyanarayan. 2019. Sociotechnical Considerations for Accessible Visualization Design. In *2019 IEEE Visualization Conference (VIS)*. 16–20. <https://doi.org/10.1109/VISUAL.2019.8933762>
- [20] Charles Perin. 2021. What Students Learn With Personal Data Physicalization. *IEEE Computer Graphics and Applications* 41, 6 (2021), 48–58. <https://doi.org/10.1109/MCG.2021.3115417>
- [21] Fabio Pittarello and Manuel Semenzato. 2024. Towards a Data Physicalization Toolkit for Non-Sighted Users. In *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*. 1–6. <https://doi.org/10.1109/CCNC51664.2024.10454861>
- [22] Kim Sauvé, Hans Brombacher, Rosa Van Koningsbruggen, Annemiek Veldhuis, Steven Houben, and Jason Alexander. 2023. Physicalization from Theory to Practice: Exploring Physicalization Design across Domains. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 360, 7 pages. <https://doi.org/10.1145/3544549.3573824>
- [23] Kim Sauvé, Pierre Dragicevic, and Yvonne Jansen. 2023. Edo: A Participatory Data Physicalization on the Climate Impact of Dietary Choices. In *Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction (Warsaw, Poland) (TEI '23)*. Association for Computing Machinery, New York, NY, USA, Article 35, 13 pages. <https://doi.org/10.1145/3569009.3572807>
- [24] Denise Schmandt-Besserat. 2013. Tokens and writing: the cognitive development. In *UT Faculty/Researcher Works*. Ex Oriente.
- [25] Robert Spence. 2014. *Information Visualization: An Introduction* (3rd ed.). Springer Publishing Company Incorporated.
- [26] Adam Strantz. 2023. Print-and-Play: Data Physicalization Methods for Research Analysis in Technical Communication. In *Proceedings of the 41st ACM International Conference on Design of Communication (Orlando, FL, USA) (SIGDOC '23)*. Association for Computing Machinery, New York, NY, USA, 215–220. <https://doi.org/10.1145/3615335.3623039>
- [27] Simon Stusak, Andreas Butz, and Aurélien Tabard. 2018. Variables for Data Physicalization Units. *Toward a Design Language for Data Physicalization, IEEE VIS*. (2018).
- [28] W3C. 2024. *Web Accessibility Initiative (WAI)*. Retrieved January 21, 2024 from <https://www.w3.org/WAI/>
- [29] Data Physicalization Wiki. 2022. *Global Cities: Elevation Maps of City Population*. Retrieved January 21, 2024 from <http://dataphys.org/list/city-populations/>
- [30] Data Physicalization Wiki. 2022. *Summer in the City*. Retrieved January 21, 2024 from <http://dataphys.org/list/summer-in-the-city/>
- [31] David Worrall. 2019. *Sonification Design: From Data to Intelligible Soundfields* (1st ed.). Springer Publishing Company, Incorporated.
- [32] Jack Zhao and Andrew Vande Moere. 2008. Embodiment in data sculpture: a model of the physical visualization of information. In *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts*. 343–350.